A Fast Distributed Stateless Algorithm for α-Fair Packing Problems

Jelena Marašević[*]
Columbia University
jelena@ee.columbia.edu

Cliff Stein [†]
Columbia University
cliff@ieor.columbia.edu

Gil Zussman[‡]
Columbia University
gil@ee.columbia.edu

Abstract

Over the past two decades, fair resource allocation problems have received considerable attention in a variety of application areas. However, *little progress has been made in the design of distributed algorithms with convergence guarantees for general and commonly used α-fair allocations*. In this paper, we study weighted α-fair packing problems, that is, the problems of maximizing the objective functions (i) $\sum_j w_j x_j^{1-\alpha}/(1-\alpha)$ when $\alpha > 0$, $\alpha \neq 1$ and (ii) $\sum_j w_j \ln x_j$ when $\alpha = 1$, over linear constraints $Ax \leq b$, $x \geq 0$, where w_j are positive weights and A and b are non-negative. We consider the distributed computation model that was used for packing linear programs and network utility maximization problems. Under this model, *we provide a distributed algorithm for general α* that converges to an $\varepsilon-$approximate solution in time (number of distributed iterations) that has an inverse polynomial dependence on the approximation parameter ε and poly-logarithmic dependence on the problem size. *This is the first distributed algorithm for weighted $\alpha-$fair packing with poly-logarithmic convergence in the input size.* The algorithm uses simple local update rules and is stateless (namely, it allows asynchronous updates, is self-stabilizing, and allows incremental and local adjustments). We also obtain a number of structural results that characterize $\alpha-$fair allocations as the value of α is varied. These results deepen our understanding of fairness guarantees in $\alpha-$fair packing allocations, and also provide insight into the behavior of $\alpha-$fair allocations in the asymptotic cases $\alpha \to 0$, $\alpha \to 1$, and $\alpha \to \infty$.

1 Introduction

Over the past two decades, *fair resource allocation* problems have received considerable attention in many application areas, including Internet congestion control [32], rate control in software defined networks [35], scheduling in wireless networks [46], multi-resource allocation and scheduling in datacenters [12,20,21,24], and a variety of applications in operations research, economics, and game theory [11,23]. In most of these applications, positive linear (packing) constraints arise as a natural model of the allowable allocations.

In this paper, we focus on the problem of finding an α-*fair* vector on the set determined by packing constraints $Ax \leq \mathbb{1}, x \geq 0$ where all $A_{ij} \geq 0$.[1] We refer to this problem as $\alpha-$fair packing. For a vector of positive weights w and $\alpha \geq 0$, an allocation vector x^* of size n is weighted α-fair, if for any alternative feasible vector x: $\sum_j w_j \frac{x_j - x_j^*}{(x_j^*)^\alpha} \leq 0$ [38]. For a compact and convex feasible region, x^* can be equivalently defined as a vector that solves the problem of maximizing $p_\alpha(x) = \sum_j w_j f_\alpha(x_j)$ [38], where:

$$f_\alpha(x_j) = \begin{cases} \ln(x_j), & \text{if } \alpha = 1 \\ \frac{x_j^{1-\alpha}}{1-\alpha}, & \text{if } \alpha \neq 1 \end{cases}. \tag{1}$$

α-fairness provides a trade-off between efficiency (sum of allocated resources) and fairness (minimum allocated resource) as a function of α: the higher the α, the better the fairness guarantees and the lower the efficiency [4,11,31]. Important special cases are proportional fairness ($\alpha = 1$) and max-min fairness ($\alpha \to \infty$). When $\alpha = 0$, we have the "unfair" case of linear optimization.

Distributed algorithms for $\alpha-$fair packing are of particular interest, as many applications are inherently distributed (such as, e.g., network congestion control), while in others parallelization is highly desirable due to the large problem size (as in, e.g., resource allocation in datacenters). We adopt the model of distributed computation commonly used in the design of packing linear programming (LP) algorithms [3,7,8,29,33,42] and which generalizes the model from network congestion control [26]. In this model, an agent j controls the variable x_j and has information about: (i) the jth column of the $m \times n$ constraint matrix A, (ii) the weight w_j, (iii) upper bounds on the global problem parameters $m, n, w_{\max}$, and $A_{\max}$, where $w_{\max} = \max_j w_j$, and $A_{\max} = \max_{ij} A_{ij}$, and (iv) in each round, the relative slack of each constraint i in which x_j takes part.

Distributed algorithms for $\alpha-$fair resource allocations have been most widely studied in the network congestion control literature, using a control-theoretic approach [25,26,32,38,41,46]. Such an approach yields continuous-time algorithms that converge after "finite" time; however, the convergence time of these algorithms as a function of the input size is poorly understood. Some other distributed pseudo-polynomial-time approximation algorithms that can address α-fair packing are described in Table 1. These algorithms all have convergence times that are at least linear in the parameters describing the problem.

No previous work has given truly fast (poly-log iterations) distributed algorithms for the general case of α-fair packing. Only for the unfair $\alpha = 0$ case (packing LPs), are such algorithms known [3,7,8,29,33,47].

Our Results. *We provide the first efficient, distributed, and stateless algorithm for weighted α-fair packing*, namely, for the problem $\max\{p_\alpha(x) : Ax \leq \mathbb{1}, x \geq 0\}$, where distributed agents update the values of x_j's asynchronously and react only to the current state of the constraints. We assume that all non-zero entries A_{ij} of matrix A satisfy $A_{ij} \geq 1$. Considering such a normalized form of the problem is without loss of generality (see Appendix A).

The approximation provided by the algorithm, to which we refer as the ε-approximation, is (i) $(1 + \varepsilon)$-multiplicative for $\alpha \neq 1$, and (ii) $W\varepsilon$-additive[2] for $\alpha = 1$, where $W = \sum_j w_j$. The main results are summarized in the following theorem, where, to unify the statement of the results, we treat α as a

---

[1]Although in the network congestion control literature the constraint matrix A is commonly assumed to be a 0-1 matrix [25, 26, 32, 38, 41, 46], important applications (such as, e.g., multi-resource allocation in datacenters) are modeled by a more general constraint matrix A with arbitrary non-negative elements [12, 20, 21, 24].

[2]Note that W cannot be avoided here, as additive approximation is not invariant to the scaling of the objective.

constant that is either equal to 1 or bounded away from 0 and 1, and we also loosen the bound in terms of $\varepsilon^{-1}, n, m, R_w = \max_{j,k} w_j/w_k$, and $A_{\max}$. For a more detailed statement, see Theorems 4.1 – 4.3.

Theorem 1.1. *(Main Result) For a given weighted α-fair packing problem $\max\{\sum_j w_j f_\alpha(x_j) : Ax \leq \mathbb{1}, x \geq 0\}$, where $f_\alpha(x_j)$ is given by (1), there exists a stateless and distributed algorithm (α-FAIRPSOLVER) that computes an ε-approximate solution in $O(\varepsilon^{-5} \ln^4(R_w n m A_{\max} \varepsilon^{-1}))$ rounds.*

To the best of our knowledge, for any constant approximation parameter ε, our algorithm is the first distributed algorithm for weighted α-fair packing problems with a poly-logarithmic convergence time.

The algorithm is *stateless* according to the definition given by Awerbuch and Khandekar [6,7]: it starts from any initial state, the agents update the variables x_j in a cooperative but uncoordinated manner, reacting only to the current state of the constraints that they observe, and without access to a global clock. Statelessness implies various desirable properties of a distributed algorithm, such as: asynchronous updates, self-stabilization, and incremental and local adjustments [6,7].

We also obtain the following structural results that characterize $\alpha-$fair packing allocations as a function of the value of α:

- We derive a lower bound on the minimum coordinate of the $\alpha-$fair packing allocation as a function of α and the problem parameters (Lemma 4.29). This bound deepens our understanding of how the fairness (a minimum allocated value) changes with α.

- We prove that for $\alpha \leq \frac{\varepsilon/4}{\ln(nA_{\max}/\varepsilon)}$, $\alpha-$fair packing can be $O(\varepsilon)-$approximated by any $\varepsilon-$approximation packing LP solver (Lemma 4.30). This result allows us to focus on the $\alpha > \frac{\varepsilon/4}{\ln(nA_{\max}/\varepsilon)}$ cases.

- We show that for $|\alpha - 1| = O(\varepsilon^2/\ln^2(\varepsilon^{-1} R_w mn A_{\max}))$, $\alpha-$fair allocation is $\varepsilon-$approximated by a $1-$fair allocation returned by our algorithm (Lemmas 4.31 and 4.32).

- We show that for $\alpha \geq \ln(R_w n A_{\max})/\varepsilon$, the $\alpha-$fair packing allocation x^* and the max-min fair allocation z^* are ε-close to each other: $(1 - \varepsilon)z^* \leq x^* \leq (1 + \varepsilon)z^*$ element-wise. This result is especially interesting as (i) max-min fair packing is not a convex problem, but rather a multi-objective problem (see, e.g., [27,44]) and (ii) the result yields the first convex relaxation of max-min fair allocation problems with a $1 \pm \varepsilon$ gap.

We now overview some of the main technical details of α-FAIRPSOLVER. In doing so, we point out connections to the two main bodies of previous work, from packing LPs [7] and network congestion control [25]. We also outline the new algorithmic ideas and proofs that were needed to obtain the results.

The algorithm and KKT conditions. The algorithm maintains primal and dual feasible solutions and updates each primal variable x_j whenever a Karush-Kuhn-Tucker (KKT) condition $x_j^\alpha \sum_i y_i A_{ij} = w_j$ is not *approximately* satisfied. In previous work, relevant update rules include: [25] (for $\alpha = 1$), where the update of each variable x_j is proportional to the difference $w_j - x_j^\alpha \sum_i y_i A_{ij}$, and [7] (for $\alpha = 0$), where each x_j is updated by a multiplicative factor $1 \pm \beta$, whenever $x_j^\alpha \sum_i y_i A_{ij} = w_j$ is not approximately satisfied. For our techniques (addressing a general α) such rules do not suffice and we introduce the following modifications: (i) in the $\alpha < 1$ case we use multiplicative updates by factors $(1 + \beta_1)$ and $(1 - \beta_2)$, where $\beta_1 \neq \beta_2$ and (ii) we use additional threshold values δ_j to make sure that x_j's do not become too small. These thresholds guarantee that we maintain a feasible solution, but they significantly complicate (compared to the linear case) the argument that each step makes a significant progress.

Dual Variables. In α-FAIRPSOLVER, a dual variable y_i is an exponential function of the ith constraint's relative slack: $y_i(x) = C \cdot e^{\kappa(\sum_j A_{ij}x_j - 1)}$, where C and κ are functions of global input parameters $\alpha, w_{\max}, n, m$, and $A_{\max}$. Packing LP algorithms [3, 7, 8, 17, 18, 28, 43] use similar dual variables with $C = 1$. Our work requires choosing C to be a function of $\alpha, w_{\max}, n, m, A_{\max}$ rather than a constant.

Paper	Number of Distributed Iterations[3]	Statelessness	Notes
[15]	$\Omega(\varepsilon^{-1} n A_{\max})$	Semi-stateless[4]	Only for $\alpha = 1$
[9]	$\Omega(\varepsilon^{-1} m n A_{\max}{}^2)$	Not stateless	
[39]	$\text{poly}(\varepsilon^{-1}, m, n, A_{\max})$	Semi-stateless	
[this work]	$O(\varepsilon^{-5} \ln^4(R_w m n A_{\max}/\varepsilon))$	Stateless	

Table 1: Comparison among distributed algorithms for $\alpha-$fair packing.

Convergence Argument. The convergence analysis of α-FAIRPSOLVER relies on the appropriately chosen concave potential function that is bounded below and above for $x_j \in [\delta_j, 1]$, $\forall j$, and that increases with every primal update. The algorithm can also be interpreted as a gradient ascent on a regularized objective function (the potential function), using a generalized entropy regularizer (see [1, 3]). A similar potential function was used in many works on packing and covering linear programs, such as, e.g., in [7] and (implicitly) in [47]. The Lyapunov function from [25] is also equivalent to this potential function when $y_i(x) = C \cdot e^{\kappa(\sum_j A_{ij} x_j - 1)}$, $\forall i$. As in these works, the main idea in the analysis is to show that whenever a solution x is not "close" to the optimal one, the potential function increases substantially. However, our work requires several new ideas in the convergence proofs, the most notable being *stationary rounds*. A stationary round is roughly a time when the variables x_j do not change much and are close to the optimum. Poly-logarithmic convergence time is then obtained by showing that: (i) there is at most a poly-logarithmic number of non-stationary rounds where the potential function increases additively and the increase is "large enough", and (ii) in all the remaining non-stationary rounds, the potential function increases multiplicatively. Our use of stationary rounds is new, as is the use of Lagrangian duality and all the arguments that follow.

Relationship to Previous Work. Very little progress has been made in the design of efficient distributed algorithms for the general class of α-fair objectives. Classical work on distributed rate control algorithms in the networking literature uses a control-theoretic approach to optimize α-fair objectives. While such an approach has been extensively studied and applied to various network settings [25, 26, 32, 38, 41, 46], it has never been proven to have polynomial convergence time (and it is unclear whether such a result can be established).

Since α-fair objectives are concave, their optimization over a region determined by linear constraints is solvable in polynomial time in a centralized setting through convex programming (see, e.g., [13, 40]). Distributed gradient methods for network utility maximization problems, such as e.g., [9, 39] summarized in Table 1, can be employed to address the problem of α-fair packing. However, the convergence times of these algorithms depend on the dual gradient's Lipschitz constant to produce good approximations. While [9, 39] provide a better dependence on the accuracy ε than our work, the dependence on the dual gradient's Lipschitz constant, in general, leads to at least linear convergence time as a function of n, m, and $A_{\max}$.

As mentioned before, some special cases have been addressed, particularly for max-min fairness ($\alpha \to \infty$) and for packing LPs ($\alpha = 0$). Relevant work on max-min fairness includes [10, 14, 22, 27, 30, 34, 36], but none of these works have poly-logarithmic convergence time. There is a long history of interesting work on packing LPs in both centralized and distributed settings, e.g., [1, 3, 7, 8, 18, 19, 28, 29, 33, 43, 47]. Only a few of these works are stateless, including the packing LP algorithm of Awerbuch and Khandekar [7], flow control algorithm of Garg and Young [19], and the algorithm of Awerbuch, Azar, and Khandekar [5] for the special case of load balancing in bipartite graphs. Additionally, the packing LP algorithm of Allen-Zhu and Orecchia [3] is "semi-stateless"; the lacking property to make it stateless is that it requires synchronous updates. The $\alpha = 1$ case of α-fair packing problems is equivalent to the problem of finding an equilibrium allocation

---

[3]The convergence times in [9, 15, 39] are not stated only in terms of the input parameters, but also in terms of intermediary parameters that depend on the problem structure. Stated here are our lowest estimates of the worst-case convergence times.

[4]A distributed algorithm is semi-stateless, if all the updates depend only on the current state of the constraints, the updates are performed in a cooperative but non-coordinated manner, and *the updates need to be synchronous* [3].

in Eisenberg-Gale markets with Leontief utilities (see [15]). Similar to the aforementioned algorithms, the algorithm from [15] converges in time linear in ε^{-1} but also (at least) linear in the input size (see Table 1).

In terms of the techniques, closest to our work is the work by Awerbuch and Khandekar [7] and we now highlight the differences compared to this work. *Some preliminaries* of the convergence proof follow closely those from [7]: mainly, Lemmas 4.4, 4.6, and 4.9 use similar arguments as corresponding lemmas in [7]. Some parts of the lemmas lower-bounding the potential increase in $\alpha < 1$, $\alpha = 1$, and $\alpha > 1$ cases (Lemmas 4.10, 4.16, and 4.22) use similar arguments as [7], however, even those parts require additional results due to the existence of lower thresholds δ_j.

The similarity ends here, as the main convergence arguments are different than those used in [7]. In particular, the convergence argument from [7] relying on stationary intervals cannot be applied in the setting of $\alpha-$fair objectives. More details about why this argument cannot be applied and where it fails are provided in Section 4. As already mentioned, we rely on the appropriately chosen definition of a stationary round. To show that in a stationary round a solution x is $\varepsilon-$approximate, we use Lagrangian duality and bound the duality gap through an intricate case analysis. We remark that such an argument could not have been used in [7], since in the packing LP case there is no guarantee that the solution y is dual-feasible.

Organization of the Paper. The rest of the paper is organized as follows. Section 2 provides the background. Section 3 describes the algorithm, and Section 4 provides the convergence analysis and structural results. Section 5 concludes the paper.

2 Preliminaries

Weighted α-Fair Packing. Consider the following optimization problem with positive linear (packing) constraints: $(Q_\alpha) = \max\{p_\alpha(x) \equiv \sum_{j=1}^{n} w_j f_\alpha(x_j) : Ax \leq b, x \geq 0\}$, where $f_\alpha(x_j)$ is given by (1), $x = (x_1, ..., x_n)$ is the vector of variables, A is an $m \times n$ matrix with non-negative elements, and $b = (b_1, ..., b_m)$ is a vector with strictly positive[5] elements. We refer to (Q_α) as the weighted α-fair packing. The following definition and lemma introduced by Mo and Walrand [38] characterize weighted α-fair allocations. In the rest of the paper, we will use the terms weighted α-fair and α-fair interchangeably.

Definition 2.1. [38] *Let $w = (w_1, ..., w_n)$ be a vector with positive entries and $\alpha > 0$. A vector $x^* = (x_1^*, ..., x_n^*)$ is weighted α-fair, if it is feasible and for any other feasible vector x: $\sum_{j=1}^{n} w_j \frac{x_j - x_j^*}{x_j^{*\alpha}} \leq 0$.*

Lemma 2.2. [38] *A vector x^* solves (Q_α) for functions $f_\alpha(x_j^*)$ if and only if it is weighted α-fair.*

Notice in (Q_α) that since $b_i > 0$, $\forall i$, and the partial derivative of the objective with respect to any of the variables x_j goes to ∞ as $x_j \to 0$, the optimal solution must lie in the positive orthant. Moreover, since the objective is strictly concave and maximized over a convex region, the optimal solution is unique and (Q_α) satisfies strong duality (see, e.g., [13]). The same observations are true for the scaled version of the problem denoted by (P_α) and introduced in the following subsection.

Normalized Form. We consider the weighted α-fair packing problem in the normalized form:

$$(P_\alpha) = \max\{p_\alpha(x) : Ax \leq \mathbb{1}, x \geq 0\},$$

where $p_\alpha(x) = \sum_{j=1}^{n} w_j f_\alpha(x_j)$, f_α is defined by (1), $w = (w_1, ..., w_n)$ is a vector of positive weights, $x = (x_1, ..., x_n)$ is the vector of variables, A is an $m \times n$ matrix with non-negative entries, and $\mathbb{1}$ is a size-m vector of 1's. We let $A_{\max}$ denote the maximum element of the constraint matrix A, and assume that every entry A_{ij} of A is non-negative, and moreover, that $A_{ij} \geq 1$ whenever $A_{ij} \neq 0$. The maximum weight is denoted by $w_{\max}$ and the minimum weight is denoted by $w_{\min}$. The sum of the weights is denoted by W and the ratio $\frac{w_{\max}}{w_{\min}}$ by R_w. We remark that considering problem (Q_α) in the normalized form (P_α) is without loss of generality: any problem (Q_α) can be scaled to this form by (i) dividing both sides of each inequality

---

[5]If, for some i, $b_i = 0$, then trivially $x_j = 0$, for all j such that $A_{ij} \neq 0$.

i by b_i and (ii) working with scaled variables $c \cdot x_j$, where $c = \min\{1, \min_{\{i,j:A_{ij}\neq0\}} \frac{A_{ij}}{b_i}\}$. Moreover, such scaling preserves the approximation (A).

KKT Conditions and Duality Gap We will denote the Lagrange multipliers for (P_α) as $y = (y_1, ..., y_m)$ and refer to them as "dual variables". The KKT conditions for (P_α) are (see Appendix B):

$$\sum_{j=1}^{n} A_{ij}x_j \leq 1, \quad \forall i \in \{1, ..., m\}; \quad x_j \geq 0, \quad \forall j \in \{1, ..., n\} \quad \text{(primal feasibility)} \tag{K1}$$

$$y_i \geq 0, \quad \forall i \in \{1, ..., m\} \quad \text{(dual feasibility)} \tag{K2}$$

$$y_i \cdot \Big(\sum_{j=1}^{m} A_{ij}x_j - 1 \Big) = 0, \quad \forall i \in \{1, ..., m\} \quad \text{(complementary slackness)} \tag{K3}$$

$$x_j{}^{\alpha} \sum_{i=1}^{m} y_i A_{ij} = w_j, \quad \forall j \in \{1, ..., m\} \quad \text{(gradient conditions)} \tag{K4}$$

The duality gap for $\alpha \neq 1$ is (see Appendix B):

$$G_\alpha(x, y) = \sum_{j=1}^{n} w_j \frac{x_j{}^{1-\alpha}}{1-\alpha}\big(\xi_j{}^{\frac{\alpha-1}{\alpha}} - 1\big) + \sum_{i=1}^{m} y_i - \sum_{j=1}^{n} w_j x_j{}^{1-\alpha} \cdot \xi_j{}^{\frac{\alpha-1}{\alpha}}, \tag{2}$$

where $\xi_j = \frac{x_j{}^{\alpha}\sum_{i=1}^{m} y_i A_{ij}}{w_j}$, while for $\alpha = 1$:

$$G_1(x, y) = -\sum_{j=1}^{n} w_j \ln\Big(\frac{x_j\sum_{i=1}^{m} y_i A_{ij}}{w_j}\Big) + \sum_{i=1}^{m} y_i - W. \tag{3}$$

Model of Distributed Computation We adopt the same model of distributed computation as [3, 7, 8, 29, 33, 42], described as follows. We assume that for each $j \in \{1, ..., n\}$, there is an agent controlling the variable x_j. Agent j is assumed to have information about the following problem parameters: (i) the jth column of A, (ii) the weight w_j, and (iii) (an upper bound on) $m, n, w_{\max}$, and $A_{\max}$. In each round, agent j collects the relative slack[6] $1 - \sum_{j=1}^{n} A_{ij}x_j$ of all constraints i for which $A_{ij} \neq 0$.

We remark that this model of distributed computation is a generalization of the model considered in network congestion control problems [26] where a variable x_j corresponds to the rate of node j, A is a 0-1 routing matrix, such that $A_{ij} = 1$ if and only if a node j sends flow over link i, and b is the vector of link capacities. Under this model, the knowledge about the relative slack of each constraint corresponds to each node collecting (a function of) congestion on each link that it utilizes. Such a model was used in network utility maximization problems with α-fair objectives [25] and general strongly-concave objectives [9].

3 Algorithm

The pseudocode for the α-FAIRPSOLVER algorithm that is run at each node j is provided in Fig 1. The basic intuition is that the algorithm keeps KKT conditions (K1) and (K2) satisfied and works towards (approximately) satisfying the remaining two KKT conditions (K3) and (K4) to minimize the duality gap. The algorithm can run in the distributed setting described in Section 2. In each round, an agent j updates the value of x_j based on the relative slack of all the constraints in which j takes part, as long as the KKT condition (K4) of agent j is not approximately satisfied. The updates need not be synchronous: we will require that all agents make updates at the same speed, but without access to a global clock.

To allow for self-stabilization and dynamic changes, the algorithm runs forever at all the agents, which is a standard requirement for self-stabilizing algorithms (see, e.g., [16]). The convergence of the algorithm is measured as the number of rounds between the round in which the algorithm starts from some initial solution and the round in which it reaches an $\varepsilon-$approximate solution, assuming that there are no hard reset events or node/constraint insertions/deletions in between.

---

[6]The slack is "relative" because in a non-scaled version of the problem where one could have $b_i \neq 1$, agent j would need to have information about $\frac{b_i-\sum_{j=1}^{n} A_{ij}x_j}{b_i}$.

α-FAIRPSOLVER(ε)

(Parameters $\delta_j, C, \kappa, \gamma, \beta_1$, and β_2 are set as described in the text below the algorithm.)
In each round of the algorithm:

1: $x_j \leftarrow \max\{x_j, \delta_j\}, x_j = \min\{x_j, 1\}$
2: Update the dual variables: $y_i = C \cdot e^{\kappa\left(\sum_{j=1}^n A_{ij}x_j - 1\right)} \ \forall i \in \{1, ..., m\}$
3: **if** $\frac{x_j{}^\alpha \cdot \sum_{i=1}^m y_i A_{ij}}{w_j} \leq (1 - \gamma)$ **then**
4: $\quad x_j \leftarrow x_j \cdot (1 + \beta_1)$
5: **else**
6: $\quad$ **if** $\frac{x_j{}^\alpha \cdot \sum_{i=1}^m y_i A_{ij}}{w_j} \geq (1 + \gamma)$ **then**
7: $\quad\quad x_j \leftarrow \max\{x_j \cdot (1 - \beta_2), \delta_j\}$

Figure 1: Pseudocode of α-FAIRPSOLVER algorithm.

Without loss of generality, we assume that the input parameter ε that determines the approximation quality satisfies $\varepsilon \leq \min\{\frac{1}{6}, \frac{9}{10\alpha}\}$ for any α, and $\varepsilon \leq \frac{1-\alpha}{\alpha}$ for $\alpha < 1$. The parameters $\delta_j, C, \kappa, \gamma, \beta_1$, and β_2 are set as follows. For technical reasons (mainly due to reinforcing dominant multiplicative updates of the variables x_j), we set the values of the lower thresholds δ_j below the actual lower bound of the optimal solution that we derive in Lemma 4.29:

$$\delta_j = \left(\frac{1}{2} \cdot \frac{w_j}{w_{\max}}\right)^{1/\alpha} \cdot \begin{cases} \left(\frac{1}{m \cdot n^2 \cdot A_{\max}}\right)^{1/\alpha}, & \text{if } 0 < \alpha \leq 1 \\ \frac{1}{m \cdot n^2 A_{\max}{}^{2-1/\alpha}}, & \text{if } \alpha > 1 \end{cases}.$$

We denote $\delta_{\max} \equiv \max_j \delta_j$, $\delta_{\min} \equiv \min_j \delta_j$. The constant C that multiplies the exponent in the dual variables y_i is chosen as $C = \frac{W}{\sum_{j=1}^n \delta_j{}^\alpha}$. Because δ_j only depends on w_j and on global parameters, we also have $C = \frac{w_j}{\delta_j{}^\alpha}, \forall j$. The parameter κ that appears in the exponent of the y_i's is chosen as $\kappa = \frac{1}{\varepsilon} \ln\left(\frac{CmA_{\max}}{\varepsilon w_{\min}}\right)$. The "absolute error" of (K4) γ is set to $\varepsilon/4$. For $\alpha \geq 1$, we set $\beta_1 = \beta_2 = \beta$, where the choice of β is described below. For $\alpha < 1$, we set $\beta_1 = \beta, \beta_2 = \beta^2(\ln(\frac{1}{\delta_{\min}}))^{-1}$.

Similar to [7], we choose the value of β so that if we set $\beta_1 = \beta_2 = \beta$, in any round the value of each $\frac{x_j{}^\alpha \sum_{i=1}^m y_i(x) A_{ij}}{w_j}$ changes by a multiplicative factor of at most $(1 \pm \gamma/4)$. Since the maximum increase over any x_j in each iteration is by a factor $1 + \beta$, and x is feasible in each round (see Lemma 4.4), we have that $\sum_{j=1}^n A_{ij}x_j \leq 1$, and therefore, the maximum increase in each y_i is by a factor of $e^{\kappa\beta}$. A similar argument holds for the maximum decrease. Hence, we choose β so that:

$$(1 + \beta)^\alpha e^{\kappa\beta} \leq 1 + \gamma/4 \quad \text{and} \quad (1 - \beta)^\alpha e^{-\kappa\beta} \geq 1 - \gamma/4,$$

and it suffices to set:

$$\beta = \begin{cases} \frac{\gamma}{5(\kappa+1)}, & \text{if } \alpha \leq 1 \\ \frac{\gamma}{5(\kappa+\alpha)}, & \text{if } \alpha > 1 \end{cases}.$$

Remark: In the $\alpha < 1$ cases, since $\beta_2 = \beta^2(\ln(1/\delta_{\min}))^{-1}$, the maximum decrease in $\frac{x_j{}^\alpha \sum_i y_i(x) A_{ij}}{w_j}$ is by a factor $(1 - (\gamma/4) \cdot \beta(\ln(1/\delta_{\min}))^{-1}), \forall j$.

4 Convergence Analysis

In this section, we analyze the convergence time of α-FAIRPSOLVER. We first state our main theorems and provide some general results that hold for all $\alpha > 0$. We show that starting from an arbitrary solution, the algorithm reaches a feasible solution within poly-logarithmic (in the input size) number of rounds, and maintains a feasible solution forever after. Similar to [7, 25, 47], we use a concave potential function that, for feasible x, is bounded below and above and increases with any algorithm update. Then, we analyze the convergence time separately for three cases: $\alpha < 1$, $\alpha = 1$, and $\alpha > 1$. With an appropriate definition of a

stationary round for each of the three cases, we show that in every stationary round, x approximates "well" the optimal solution by bounding the duality gap. On the other hand, for any non-stationary round, we show that the potential increases substantially. This large increase in the potential then leads to the conclusion that there cannot be too many non-stationary rounds, thus bounding the overall convergence time.

We make a few remarks here. First, we require that α be bounded away from zero. This requirement is without loss of generality because we show that when $\alpha \leq \frac{\varepsilon/4}{\ln(nA_{\max}/\varepsilon)}$, any $\varepsilon-$approximation LP provides a $3\varepsilon-$approximate solution to (P_α) (Lemma 4.30). Thus, when $\alpha \leq \frac{\varepsilon/4}{\ln(nA_{\max}/\varepsilon)}$ we can switch to the algorithm of [7], and when $\alpha > \frac{\varepsilon/4}{\ln(nA_{\max}/\varepsilon)}$, the convergence time remains poly-logarithmic in the input size and polynomial in ε^{-1}. Second, the assumption that $\varepsilon \leq \frac{1-\alpha}{\alpha}$ in the $\alpha < 1$ case is also without loss of generality, because we show that when α is close to 1 (roughly, $1 - O(\varepsilon^2/\ln^2(R_w mnA_{\max}/\varepsilon))$), we can approximate (P_α) by switching to the $\alpha = 1$ case of the algorithm (Lemma 4.31). Finally, when $\alpha > 1$, the algorithm achieves an $\varepsilon-$approximation in time $O(\alpha^4\varepsilon^{-4}\ln^2(R_w nmA_{\max}\varepsilon^{-1}))$. We believe that a polynomial dependence on α is difficult to avoid in this setting, because by increasing α, the gradient of the α-fair utilities f_α blows up on the interval $(0,1)$: as α increases, $f_\alpha(x)$ quickly starts approaching a step function that is equal to $-\infty$ on the interval $(0,1]$ and equal to 0 on the interval $(1,\infty]$. To characterize the behavior of $\alpha-$fair allocations as α becomes large, we show that when $\alpha \geq \varepsilon^{-1}\ln(R_w nA_{\max})$, all the coordinates of the $\alpha-$fair vector are within a $1 \pm \varepsilon$ multiplicative factor of the corresponding coordinates of the max-min fair vector (Lemma 4.34).

Finally, we note that the main convergence argument from [7] that uses an appropriate definition of *stationary intervals* does not extend to our setting. The proof from [7] "breaks" in the part that shows that the solution is $\varepsilon-$approximate throughout any stationary interval, stated as Lemma 3.7 in [7]. The proof of Lemma 3.7 in [7] is by contradiction: assuming that the solution is not $\varepsilon-$approximate, the proof proceeds by showing that at least one of the variables would increase in each round of the stationary interval, thus eventually making the solution infeasible and contradicting one of the preliminary lemmas. For $\alpha \geq 1$, unlike the linear objective in [7], α-fair objectives are negative, and the assumption that the solution is not $\varepsilon-$approximate does not lead to any conclusive information. For $\alpha < 1$, adapting the proof of Lemma 3.7 from [7] leads to the conclusion that for at least one j, in each round t of the stationary interval $\frac{(x_j^*)^\alpha \sum_i y_i(x^t)A_{ij}}{w_j} \leq 1 - \gamma$, where x^* is the optimal solution, and x^t is the solution at round t. In [7], where $\alpha = 0$, this implies that x_j increases in each round of the stationary interval, while in our setting ($\alpha > 0$) it is not possible to draw such a conclusion.

Main Results. Our main results are summarized in the following three theorems. The objective is denoted by $p_\alpha(x)$, x^t denotes the solution at the beginning of round t, and x^* denotes the optimal solution.

Theorem 4.1. *(Convergence for $\alpha < 1$)* α-FAIRPSOLVER *solves* (P_α) *approximately for $\alpha < 1$ in time that is polynomial in* $\frac{\ln(nmA_{\max})}{\alpha\varepsilon}$. *In particular, after at most*

$$O\left(\alpha^{-2}\varepsilon^{-5}\ln^2(R_w mnA_{\max})\ln^2\left(\varepsilon^{-1}R_w mnA_{\max}\right)\right) \tag{4}$$

rounds, there exists at least one round t such that $p_\alpha(x^*) - p_\alpha(x^t) \leq \varepsilon p_\alpha(x^t)$. Moreover, the total number of rounds s in which $p_\alpha(x^*) - p_\alpha(x^s) > \varepsilon p_\alpha(x^s)$ is also bounded by (4).

Theorem 4.2. *(Convergence for $\alpha = 1$)* α-FAIRPSOLVER *solves* (P_1) *approximately in time that is polynomial in* $\varepsilon^{-1}\ln(R_w nmA_{\max})$. *In particular, after at most*

$$O\left(\varepsilon^{-5}\ln^2(R_w nmA_{\max})\ln^2\left(\varepsilon^{-1}R_w nmA_{\max}\right)\right) \tag{5}$$

rounds, there exists at least one round t such that $p(x^*) - p(x^t) \leq \varepsilon W$. Moreover, the total number of rounds s in which $p(x^*) - p(x^s) > \varepsilon W$ is also bounded by (5).

Theorem 4.3. *(Convergence for $\alpha > 1$)* α-FAIRPSOLVER *solves* (P_α) *approximately for $\alpha > 1$ in time that*

is polynomial in $\varepsilon^{-1} \ln(nmA_{\max})$. In particular, after at most:

$$O\left(\alpha^4 \varepsilon^{-4} \ln\left(R_w nmA_{\max}\right) \ln\left(\varepsilon^{-1} R_w nmA_{\max}\right)\right) \tag{6}$$

rounds, there exists at least one round t such that $p_\alpha(x^*) - p_\alpha(x^t) \leq \varepsilon(-p_\alpha(x^t))$. Moreover, the total number of rounds s in which $p_\alpha(x^*) - p_\alpha(x^s) > \varepsilon(-p_\alpha(x^s))$ is also bounded by (6).

Feasibility and Approximate Complementary Slackness. The following three lemmas are preliminaries for the convergence time analysis. Lemma 4.4 shows that starting from a feasible solution, the algorithm always maintains a feasible solution. Lemma 4.5 shows that any violated constraint becomes feasible within poly-logarithmic number of rounds, and remains feasible forever after. Combined with Lemma 4.4, Lemma 4.5 allows us to focus only on the rounds with feasible solutions x. Lemma 4.6 shows that after a poly-logarithmic number of rounds, approximate complementary slackness (KKT condition (K3)) holds in an aggregate sense: $\sum_{i=1}^{m} y_i(x)\left(\sum_{j=1}^{n} A_{ij}x_j - 1\right) \approx 0$.

Lemma 4.4. *If the algorithm starts from a feasible solution, then the algorithm maintains a feasible solution x: $x_j \geq 0$, $\forall j$ and $\sum_{j=1}^{n} A_{ij}x_j \leq 1$, $\forall i$, in each round.*

Proof. By the statement of the lemma, the solution is feasible initially. From the way that the algorithm makes updates to the variables x_j, it is always true that $x_j \geq 0$, $\forall j$.

Now assume that x becomes infeasible in some round, and let x^0 denote the (feasible) solution before that round, x^1 denote the (infeasible) solution after the round. We have:

$$\sum_{\ell=1}^{n} A_{i\ell}x_\ell^0 \leq 1, \quad \forall i \in \{1, ..., m\}, \quad \text{and} \quad \sum_{\ell=1}^{n} A_{k\ell}x_\ell^1 > 1, \quad \text{for some } k \in \{1, ..., m\}.$$

For this to be true, x must have increased over at least one coordinate j such that $A_{kj} \neq 0$. For such a change to be triggered by the algorithm, it must also be true that:

$$(x_j^0)^\alpha \sum_{i=1}^{m} y_i(x^0)A_{ij} \leq w_j\left(1 - \gamma\right).$$

Since, by the choice of $\beta_1 = \beta$, this term can increase by a factor of at most $1 + \gamma/4$, it follows that:

$$(x_j^1)^\alpha \sum_{i=1}^{m} y_i(x^1)A_{ij} \leq w_j(1 - \gamma)\left(1 + \frac{\gamma}{4}\right) < w_j.$$

This further implies:

$$(x_j^1)^\alpha y_k(x^1)A_{kj} < w_j,$$

and since whenever $A_{kj} \neq 0$ we also have $A_{kj} \geq 1$, we get:

$$(x_j^1)^\alpha y_k(x^1) < w_j. \tag{7}$$

On the other hand, since $x_j^1 \geq \delta_j$, $\delta_j{}^\alpha = \frac{w_j}{C}$, and $\sum_{j=1}^{n} A_{kj}x_j^1 > 1$:

$$(x_j^1)^\alpha y_k(x^1) \geq \frac{w_j}{C} \cdot C \cdot e^{\kappa(\sum_{j=1}^{n} A_{kj}x_j^1 - 1)} > w_j,$$

which contradicts (7). $\qquad\square$

Lemma 4.5. *If for any i: $\sum_{j=1}^{n} A_{ij}x_j > 1$, then after at most $\tau_1 = O(\frac{1}{\beta_2} \ln(nA_{\max}))$ rounds, it is always true that $\sum_{j=1}^{n} A_{ij}x_j \leq 1$.*

Proof. Suppose that $\sum_{j=1}^{n} A_{ij}x_j > 1$ for some i. Then $y_i > C$, and for every x_j with $A_{ij} \neq 0$:

$$x_j{}^\alpha \sum_{l=1}^{m} y_l(x)A_{lj} \geq x_j{}^\alpha y_i(x)A_{ij} \geq \delta_j{}^\alpha C \geq w_j > w_j(1 - \gamma),$$

and therefore, none of the variables that appear in i increases.

8

Since $\sum_{j=1}^{n} A_{ij}x_j > 1$, there exists at least one x_k with $A_{ik} \neq 0$ such that $x_k \geq \frac{\sum_{j=1}^{n} A_{ij}x_j}{A_{ik}n} > \frac{1}{nA_{\max}}$. For each such x_k, since $C \geq 2w_{\max}nA_{\max}$:

$$x_k{}^\alpha \sum_{l=1}^{m} y_l(x)A_{lj} \geq C\frac{1}{nA_{\max}} \geq 2w_{\max} > w_k(1+\gamma),$$

and therefore, x_k decreases (by a factor $(1-\beta_2)$). As $x_k \leq 1$, after at most $O(\frac{1}{\beta_2}\ln(nA_{\max}))$ rounds in which $\sum_{j=1}^{n} A_{ij}x_j > 1$, we must have $x_k \leq \frac{1}{nA_{\max}}$, and therefore, $\sum_{j=1}^{n} A_{ij}x_j \leq 1$.

Using the same arguments as in the proof of Lemma 4.4, the constraint i never gets violated again. $\qquad\square$

Lemma 4.6. *If the algorithm starts from a feasible solution, then after at most $\tau_0 = \frac{1}{\beta}\ln\left(\frac{1}{\delta_{\min}}\right)$ rounds, it is always true that:*

1. *There exists at least one approximately tight constraint:* $\max_i\left\{\sum_{j=1}^{n} A_{ij}x_j\right\} \geq 1 - (1+1/\kappa)\varepsilon$,
2. $\sum_{i=1}^{m} y_i \leq (1+3\varepsilon)\sum_{j=1}^{n} x_j \sum_{i=1}^{m} y_i A_{ij}$, *and*
3. $(1-3\varepsilon)\sum_{i=1}^{m} y_i \leq \sum_{j=1}^{n} x_j \sum_{i=1}^{m} y_i A_{ij} \leq \sum_{i=1}^{m} y_i$.

Proof. Suppose that $\max_i \sum_{j=1}^{n} A_{ij}x_j < 1 - \varepsilon$. Then for each y_i we have:

$$y_i \leq C \cdot e^{-\kappa\varepsilon} = C \cdot \frac{\varepsilon w_{\min}}{CmA_{\max}} = \frac{\varepsilon w_{\min}}{mA_{\max}}.$$

Due to Lemma 4.4, we have that x is feasible in every round, which implies that $x_j \leq 1 \; \forall j$. This further gives:

$$x_j{}^\alpha \sum_{i=1}^{m} y_i A_{ij} \leq w_j\varepsilon \leq w_j(1-\gamma),$$

and, therefore, all variables x_j increase by a factor $1+\beta$. From Lemma 4.4, since the solution always remains feasible, none of the variables can increase to a value larger than 1. Therefore, after at most $\tau_0 = \log_{1+\beta}\left(\frac{1}{\delta_{\max}}\right) \leq \frac{1}{\beta}\ln\left(\frac{1}{\delta_{\max}}\right)$ rounds, there must exist at least one i such that $\sum_{j=1}^{n} A_{ij}x_j \geq 1-\varepsilon$. If in any round $\max_i \sum_{j=1}^{n} A_{ij}x_j$ decreases, it can decrease by at most $\beta_2 \sum_{j=1}^{n} A_{ij}x_j \leq \beta \sum_{j=1}^{n} A_{ij}x_j \leq \beta < \frac{\varepsilon}{5\kappa}$. Therefore, in every subsequent round

$$\max_i \sum_{j=1}^{n} A_{ij}x_j > 1 - \left(1 + \frac{1}{5\kappa}\right)\varepsilon.$$

For the second part of the lemma, let $S = \{i : \sum_{j=1}^{n} A_{ij}x_j < \max_{k\in\{1,\dots,m\}} \sum_{j=1}^{n} A_{kj}x_j - \frac{\kappa-1}{5\kappa}\varepsilon\}$ be the set of constraints that are at least "$\frac{\kappa-1}{5\kappa}\varepsilon$-looser" than the tightest constraint. Then for $i \in S$ we have

$$y_i \leq e^{-\frac{\kappa-1}{5}\varepsilon} \max_{k\in\{1,\dots,m\}} y_k < \frac{\varepsilon}{m}e^{\varepsilon/5} \max_{k\in\{1,\dots,m\}} y_k < 1.2\frac{\varepsilon}{m} \max_{k\in\{1,\dots,m\}} y_k.$$

This further gives:

$$\sum_{i=1}^{m} y_i = \sum_{i\in S} y_i + \sum_{k\notin S} y_k < (1+1.2\varepsilon)\sum_{i\notin S} y_i.$$

Moreover, for each $i \notin S$ we have $y_i \sum_{j=1}^{n} A_{ij}x_j \geq (1-1.2\varepsilon)y_i$, since for $i \notin S$:

$$\sum_{j=1}^{n} A_{ij}x_j \geq \max_{k\in\{1,\dots,m\}} A_{kj}x_j - \frac{\kappa-1}{5\kappa}\varepsilon \geq 1 - \left(1 + \frac{1}{5\kappa} + \frac{\kappa-1}{5\kappa}\right)\varepsilon = 1 - 1.2\varepsilon.$$

9

Therefore:

$$\sum_{i=1}^{m} y_i < \frac{1 + 1.2\varepsilon}{1 - 1.2\varepsilon} \sum_{i \notin S} y_i \sum_{j=1}^{n} A_{ij} x_j$$

$$\leq (1 + 3\varepsilon) \sum_{i \notin S} y_i \sum_{j=1}^{n} A_{ij} x_j \quad \text{(from } \varepsilon \leq 1/6)$$

$$\leq (1 + 3\varepsilon) \sum_{i=1}^{m} y_i \sum_{j=1}^{n} A_{ij} x_j.$$

Interchanging the order of summation in the last line, we reach the desired inequality.

The proof of the last part of the lemma follows from feasibility: $\sum_j A_{ij} x_j \leq 1$, $\forall i$ (Lemma 4.4), and from $\frac{1}{1+3\varepsilon} \geq 1 - 3\varepsilon$. $\qquad\square$

Lemmas analogous to 4.4 and 4.6 also appear in [7]. However, the proofs of Lemmas 4.4 and 4.6 require new ideas compared to the proofs of the corresponding lemmas in [7]. We need to be much more careful in our choice of lower thresholds δ_j and constant C in the dual variables, particularly by choosing C as a function of several variables, rather than as a constant. The choice of δ_j's is also sensitive as smaller δ_j's would make the potential function range too large, while larger δ_j's would cause more frequent decrease of "small" variables. In either case, the convergence time would increase.

Decrease of Small Variables. The following lemma is also needed for the convergence analysis. It shows that if some variable x_j decreases by less than a multiplicative factor $(1 - \beta_2)$, i.e., $x_j < \frac{\delta_j}{1-\beta_2}$ and x_j decreases, then x_j must be part of at least one approximately tight constraint. This lemma will be used later to show that in any round the increase in the potential due to the decrease of "small" variables is dominated by the decrease of "large" variables (i.e., the variables that decrease by a multiplicative factor $(1 - \beta_2)$).

Lemma 4.7. *Consider the rounds that happen after the initial* $\tau_1 = O(\frac{1}{\beta_2} \ln(n A_{\max}))$ *rounds. If in some round there is a variable* $x_j < \frac{\delta_j}{1-\beta_2}$ *that decreases, then in the same round for some i with $A_{ij} \neq 0$ it holds that:* $y_i(x) \geq \frac{\sum_{l=1}^{m} A_{lj} y_l(x)}{m A_{\max}}$ *and* $\sum_{k=1}^{n} A_{ik} x_k > 1 - \frac{\varepsilon}{2}$.

Proof. Suppose that some $x_j < \frac{\delta_j}{1-\beta_2}$ triggers a decrease over the jth coordinate. The first part of the Lemma is easy to show, simply by using the argument that at least one term of a summation must be higher than the average, i.e., there exists at least one i with $A_{ij} \neq 0$ such that:

$$y_i(x) A_{ij} \geq \frac{\sum_{l=1}^{m} A_{lj} y_l(x)}{m} \quad \Rightarrow \quad y_i \geq \frac{\sum_{l=1}^{m} A_{lj} y_l(x)}{m A_{\max}}.$$

For the second part, as $x_j < \frac{\delta_j}{1-\beta_2}$, we have that:

$$x_j^{\alpha} y_i(x) \geq \frac{x_j^{\alpha} \sum_{l=1}^{m} A_{lj} y_l(x)}{m A_{\max}} \quad \Rightarrow \quad y_i(x) > \frac{(1 - \beta_2)^{\alpha}}{\delta_j^{\alpha}} \frac{x_j^{\alpha} \sum_{l=1}^{m} A_{lj} y_l(x)}{m A_{\max}}.$$

Since x_j decreases, we have that $x_j^{\alpha} \sum_{l=1}^{m} y_i(x) A_{lj} \geq w_j(1+\gamma)$, and therefore $y_i(x) > \frac{w_j}{\delta_j^{\alpha}} \frac{(1+\gamma)(1-\beta_2)^{\alpha}}{m A_{\max}}$. Moreover, as $y_i(x) = C \cdot e^{\kappa(\sum_{k=1}^{n} A_{ik} x_k - 1)}$, and $C = \frac{w_j}{\delta_j^{\alpha}}$, it follows that:

$$e^{\kappa(\sum_{k=1}^{n} A_{ik} x_k - 1)} > \frac{(1+\gamma)(1-\beta_2)^{\alpha}}{m A_{\max}}. \tag{8}$$

Observe that for $\alpha \leq 1$:

$$(1+\gamma)(1-\beta_2)^{\alpha} \geq (1+\gamma)(1-\beta_2) > \left(1 + \frac{\varepsilon}{4}\right)\left(1 - \frac{\varepsilon}{20(\kappa+1)}\right) > 1 > \sqrt{\varepsilon}, \tag{9}$$

while for $\alpha > 1$, since $\varepsilon\alpha \leq \frac{9}{10}$:

$$(1+\gamma)(1-\beta_2)^\alpha \geq (1+\gamma)(1-\alpha\beta_2) \geq (1+\gamma)\left(1-\frac{\gamma\varepsilon\alpha}{5}\right) \geq 1 > \sqrt{\varepsilon}, \tag{10}$$

where we have used the generalized Bernoulli's inequality for $(1-\beta_2)^\alpha \geq (1-\alpha\beta_2)$ [37], and then $\beta_2 = \beta = \frac{\gamma}{5(\kappa+\alpha)} < \frac{\gamma\varepsilon}{5}$. Recalling that $\kappa = \frac{1}{\varepsilon}\ln\left(\frac{CmA_{\max}}{\varepsilon w_{\min}}\right)$, and combining (8) with (9) and (10):

$$\left(\frac{\varepsilon w_{\min}}{CmA_{\max}}\right)^{\frac{1-\sum_{k=1}^n A_{ik}x_k}{\varepsilon}} > \frac{\sqrt{\varepsilon}}{mA_{\max}}.$$

Finally, as $C \geq 2w_{\max}nmA_{\max}$, it follows that $\frac{w_{\min}\varepsilon}{CmA_{\max}} \leq \frac{\varepsilon w_{\min}}{2w_{\max}nm^2A_{\max}^2} < \left(\frac{\sqrt{\varepsilon}}{mA_{\max}}\right)^2 < 1$, which gives:

$$\frac{1-\sum_{k=1}^n A_{ik}x_k}{\varepsilon} < \frac{1}{2} \Leftrightarrow \sum_{k=1}^n A_{ik}x_k > 1 - \frac{\varepsilon}{2}.$$

$\square$

Potential. We use the following potential function to analyze the convergence time:

$$\Phi(x) = p_\alpha(x) - \frac{1}{\kappa}\sum_{i=1}^m y_i(x),$$

where $p_\alpha(x) = \sum_{j=1}^n w_j f_\alpha(x_j)$ and f_α is defined by (1). The potential function is strictly concave and its partial derivative with respect to any variable x_j is:

$$\frac{\partial\Phi(x)}{\partial x_j} = \frac{w_j}{x_j^\alpha} - \sum_{i=1}^m y_i(x)A_{ij} = \frac{w_j}{x_j^\alpha}\left(1 - \frac{x_j^\alpha\sum_{i=1}^m y_i(x)A_{ij}}{w_j}\right). \tag{11}$$

The following fact (given in a similar form in [7]), which follows directly from the Taylor series representation of concave functions, will be useful for the potential increase analysis:

Fact 4.8. *For a differentiable concave function $f : \mathbb{R}^n \to \mathbb{R}$ and any two points $x^0, x^1 \in \mathbb{R}^n$:*

$$\sum_{j=1}^n \frac{\partial f(x^0)}{\partial x_j}(x_j^1 - x_j^0) \geq f(x^1) - f(x^0) \geq \sum_{j=1}^n \frac{\partial f(x^1)}{\partial x_j}(x_j^1 - x_j^0).$$

Using Fact 4.8 and (11), we show the following lemma:

Lemma 4.9. *Starting with a feasible solution and throughout the course of the algorithm, the potential function $\Phi(x)$ never decreases. Letting x^0 and x^1 denote the values of x before and after a round update, respectively, the potential function increase is lower-bounded as:*

$$\Phi(x^1) - \Phi(x^0) \geq \sum_{j=1}^n w_j\frac{|x_j^1 - x_j^0|}{(x_j^1)^\alpha}\left|1 - \frac{(x_j^1)^\alpha\sum_{i=1}^m y_i(x^1)A_{ij}}{w_j}\right|.$$

Proof. Since Φ is concave, using Fact 4.8 and (11) it follows that:

$$\Phi(x^1) - \Phi(x^0) \geq \sum_{j=1}^n w_j\frac{x_j^1 - x_j^0}{(x_j^1)^\alpha}\left(1 - \frac{(x_j^1)^\alpha\sum_{i=1}^m y_i(x^1)A_{ij}}{w_j}\right). \tag{12}$$

If $x_j^1 = x_j^0$, then the term in the summation (12) corresponding to the change in x_j is equal to zero, and x_j has no contribution to the sum in (12).

If $x_j^1 - x_j^0 > 0$, then, as x_j increases over the observed round, it must be $\frac{(x_j^0)^\alpha\sum_{i=1}^m y_i(x^0)A_{ij}}{w_j} \leq 1 - \gamma$.

By the choice of the parameters, $\frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j} \leq \left(1 + \frac{\gamma}{4}\right)\left(\frac{(x_j^0)^\alpha \sum_{i=1}^m y_i(x^0)A_{ij}}{w_j}\right)$, and therefore

$$\frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j} \leq \left(1 + \frac{\gamma}{4}\right)(1 - \gamma) = 1 - \frac{3}{4}\gamma - \frac{\gamma^2}{4} < 1 - \frac{3}{4}\gamma. \tag{13}$$

It follows that $1 - \frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j} > \frac{3}{4}\gamma > 0$, and therefore

$$w_j \frac{x_j^1 - x_j^0}{(x_j^1)^\alpha}\left(1 - \frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j}\right) = w_j \frac{|x_j^1 - x_j^0|}{(x_j^1)^\alpha}\left|1 - \frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j}\right|.$$

Finally, if $x_j^1 - x_j^0 < 0$, then it must be $\frac{(x_j^0)^\alpha \sum_{i=1}^m y_i(x^0)A_{ij}}{w_j} \geq 1 + \gamma$. By the choice of the parameters, $\frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j} \geq \left(1 - \frac{\gamma}{4}\right)\left(\frac{(x_j^0)^\alpha \sum_{i=1}^m y_i(x^0)A_{ij}}{w_j}\right)$, implying

$$\frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j} \geq \left(1 - \frac{\gamma}{4}\right)(1 + \gamma) = 1 + \frac{3}{4}\gamma - \frac{\gamma^2}{4} > 1 + \frac{1}{2}\gamma. \tag{14}$$

We get that $1 - \frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j} < -\frac{1}{2}\gamma < 0$, and therefore

$$w_j \frac{x_j^1 - x_j^0}{(x_j^1)^\alpha}\left(1 - \frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j}\right) = w_j \frac{|x_j^1 - x_j^0|}{(x_j^1)^\alpha}\left|1 - \frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j}\right|,$$

completing the proof. $\qquad\square$

4.1 Proof of Theorem 4.1

The outline of the proof is as follows. We first derive a lower bound on the potential increase (Lemma 4.10), which will motivate the definition of a stationary round. Then, for the appropriate definition of a stationary round we will first show that in any stationary round, solution is $O(\varepsilon)-$approximate. Then, to complete the proof, we will show in any non-stationary round there is a sufficiently large increase in the potential function, which, combined with the bounds on the potential value will yield the result.

The following lemma lower-bounds the increase in the potential function in any round of the algorithm.

Lemma 4.10. *If $\alpha < 1$ and $\Phi(x^0)$, x^0, $y(x^0)$ and $\Phi(x^1)$, x^1, $y(x^1)$ denote the values of Φ, x, and y before and after a round, respectively, and $S^- = \{j : x_j \text{ decreases}\}$, then if x^0 is feasible:*

1. $\Phi(x^1) - \Phi(x^0) \geq \Omega(\beta^2 \gamma / \ln(1/\delta_{\min})) \sum_{j \in S^-} w_j \frac{(x_j^0)^{1-\alpha}}{1-\alpha}$;

2. $\Phi(x^1) - \Phi(x^0) \geq \Omega(\beta)\left((1 - \gamma)\sum_{j=1}^n w_j(x_j^0)^{1-\alpha} - \sum_{i=1}^m y_i(x^0)\sum_{j=1}^n A_{ij}x_j^0\right)$;

3. $\Phi(x^1) - \Phi(x^0) \geq \Omega\left(\frac{\beta^2}{\ln(1/\delta_{\min})}\right)\left(\sum_{i=1}^m y_i(x^0)\sum_{j=1}^n A_{ij}x_j^0 - (1 + \gamma)\sum_{j=1}^n w_j(x_j^0)^{1-\alpha}\right)$.

Proof.
Proof of 1. Observe that for $j \in S^-$, $x_j^1 = \max\{\delta_j, (1 - \beta_2)x_j^0\}$. From the proof of Lemma 4.9, we have that:

$$\Phi(x^1) - \Phi(x^0) \geq \sum_{j \in S^-} w_j \frac{x_j^0 - x_j^1}{(x_j^1)^\alpha}\left(\frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j} - 1\right).$$

The proof that

$$\sum_{j \in S^-} w_j(x_j^0)^{1-\alpha}\left(\frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j} - 1\right) = \Theta\left(\sum_{\{j \in S^- : x_j^0 \geq \frac{\delta_j}{1-\beta_2}\}} w_j(x_j^0)^{1-\alpha}\left(\frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j} - 1\right)\right)$$

12

is implied by the proof of part 3 of this lemma (see below). For each $j \in S^-$, we have that:

$$\left(\frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1) A_{ij}}{w_j} - 1 \right) \geq (1+\gamma)(1-\gamma/4) - 1 > \gamma/2,$$

Therefore:

$$\Phi(x^1) - \Phi(x^0) \geq \Omega(\gamma) \sum_{j \in S^-} w_j \frac{\beta_2 x_j^0}{(1-\beta_2)(x_j^0)^\alpha}$$

$$= \Omega\left(\frac{\beta_2 \gamma}{1-\beta_2}\right) \sum_{j \in S^-} w_j (x_j^0)^{1-\alpha}.$$

Proof of 2. Let S^+ denote the set of j's such that x_j increases in the current round. Then, recalling that for $j \in S^+$ $\frac{(x_j^0)^\alpha \sum_{i=1}^m y_i(x^0) A_{ij}}{w_j} \leq 1 - \gamma$ and that from the choice of parameters $\frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1) A_{ij}}{w_j} \leq (1 + \gamma/4) \frac{(x_j^0)^\alpha \sum_{i=1}^m y_i(x^0) A_{ij}}{w_j}$:

$$\Phi(x^1) - \Phi(x^0) \geq \sum_{j=1}^n w_j \frac{x_j^1 - x_j^0}{(x_j^1)^\alpha} \left(1 - \frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1) A_{ij}}{w_j} \right)$$

$$\geq \sum_{j \in S^+} w_j \frac{x_j^1 - x_j^0}{(x_j^1)^\alpha} \left(1 - \frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1) A_{ij}}{w_j} \right)$$

$$\geq \sum_{j \in S^+} w_j \frac{x_j^1 - x_j^0}{(x_j^1)^\alpha} \left(1 - (1 + \gamma/4) \frac{(x_j^0)^\alpha \sum_{i=1}^m y_i(x^0) A_{ij}}{w_j} \right)$$

$$\geq \sum_{j \in S^+} w_j \frac{x_j^1 - x_j^0}{(x_j^1)^\alpha} \left((1 - \gamma) - \frac{(x_j^0)^\alpha \sum_{i=1}^m y_i(x^0) A_{ij}}{w_j} \right).$$

Since $j \in S^+$, $x_j^1 = (1 + \beta) x_j^0$, it follows that

$$\Phi(x^1) - \Phi(x^0) \geq \frac{\beta}{(1+\beta)^\alpha} \sum_{j \in S^+} w_j (x_j^0)^{-\alpha} \left((1 - \gamma) - \frac{(x_j^0)^\alpha \sum_{i=1}^m y_i(x^0) A_{ij}}{w_j} \right).$$

Observing that for any $x_j \notin S^+$ we have that $(1-\gamma) - \frac{(x_j^0)^\alpha \sum_{i=1}^m y_i(x^0) A_{ij}}{w_j} < 0$, we get:

$$\Phi(x^1) - \Phi(x^0) \geq \frac{\beta}{(1+\beta)^\alpha} \sum_{j=1}^n w_j (x_j^0)^{1-\alpha} \left((1 - \gamma) - \frac{(x_j^0)^\alpha \sum_{i=1}^m y_i(x^0) A_{ij}}{w_j} \right)$$

$$= \Omega(\beta) \left((1 - \gamma) \sum_{j=1}^n w_j (x_j^0)^{1-\alpha} - \sum_{j=1}^n x_j^0 \sum_{i=1}^m y_i(x^0) A_{ij} \right).$$

Proof of 3. Let S^- denote the set of j's such that x_j decreases in the current round. In this case not all the x_j's with $j \in S^-$ decrease by a multiplicative factor $(1-\beta_2)$, since for $j \in S^-$: $x_j^1 = \max\{(1-\beta_2)x_j^0, \delta_j\}$. We will first lower-bound the potential increase over x_j's that decrease multiplicatively: $\{j : j \in S^- \land x_j^0(1 - \beta_2) \geq \delta_j\}$, so that $x_j^1 = x_j^0(1 - \beta_2)$. Recall that for $j \in S^-$: $\frac{(x_j^0)^\alpha \sum_{i=1}^m y_i(x^0) A_{ij}}{w_j} \geq 1 + \gamma$ and $\frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1) A_{ij}}{w_j} \geq (1 - \gamma/4 \frac{\beta}{\ln(1/\delta_{\min})}) \frac{(x_j^0)^\alpha \sum_{i=1}^m y_i(x^0) A_{ij}}{w_j} \geq (1 - \gamma/4) \frac{(x_j^0)^\alpha \sum_{i=1}^m y_i(x^0) A_{ij}}{w_j}$. It follows that:

$$\Phi(x^1) - \Phi(x^0) \geq \frac{\beta_2}{(1-\beta_2)^\alpha} \sum_{\{j : j \in S^- \land x_j^0(1-\beta) \geq \delta_j\}} w_j (x_j^0)^{1-\alpha} \left(\frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1) A_{ij}}{w_j} - 1 \right)$$

13

$$\geq \beta_2 \sum_{\{j:j\in S^-\wedge x_j^0(1-\beta_2)\geq \delta_j\}} w_j(x_j^0)^{1-\alpha}\left((1-\gamma/4)\frac{(x_j^0)^\alpha \sum_{i=1}^m y_i(x^0)A_{ij}}{w_j}-1\right)$$

$$= \Omega\left(\frac{\beta^2}{\ln(1/\delta_{\min})}\right)\sum_{\{j:j\in S^-\wedge x_j^0(1-\beta_2)\geq \delta_j\}} w_j(x_j^0)^{1-\alpha}\left(\frac{(x_j^0)^\alpha \sum_{i=1}^m y_i(x^0)A_{ij}}{w_j}-(1+\gamma)\right). \tag{15}$$

Next, we prove that the potential increase due to decrease of x_j such that $\{j : j \in S^- \wedge x_j^0(1-\beta_2) < \delta_j\}$ is dominated by the potential increase due to x_k's that decrease multiplicatively by the factor $(1-\beta_2)$.

Choose any x_j such that $\{j : j \in S^- \wedge x_j^0(1-\beta_2) < \delta_j\}$, and let $\xi_j(x^0) = \frac{(x_j^0)^\alpha \sum_{l=1}^m A_{lj}y_i(x^0)}{w_j}$. From Lemma 4.7, there exists at least one i with $A_{ij} \neq 0$, such that:

$$y_i \geq \frac{w_j(x_j^0)^\alpha}{w_j(x_j^0)^\alpha}\cdot\frac{\sum_{i=1}^m y_i(x^0)A_{ij}}{mA_{\max}} > \frac{1}{mA_{\max}}\frac{w_j(1-\beta_2)^\alpha}{\delta_j{}^\alpha}\xi_j(x^0) \geq \frac{1-\beta_2}{mA_{\max}}\frac{w_j}{\delta_j{}^\alpha}\xi_j(x^0), \quad \text{and,} \tag{16}$$

$$\sum_{k=1}^n A_{ik}x_k^0 > 1 - \frac{\varepsilon}{2}. \tag{17}$$

From (17), there exists at least one p such that $A_{ip} \neq 0$ and

$$A_{ip}x_p^0 > \frac{1-\frac{\varepsilon}{2}}{n}. \tag{18}$$

Since $x_p^0 \in (0,1]$ and $\alpha \in (0,1)$, using (18), we have that $A_{ip}(x_p^0)^\alpha \geq A_{ip}x_p^0 > \frac{1-\frac{\varepsilon}{2}}{n}$. Recalling (16):

$$(x_p^0)^\alpha \sum_{l=1}^m A_{lp}y_l(x^0) \geq (x_p^0)^\alpha A_{ip}y_i(x^0)$$

$$\geq \frac{1-\frac{\varepsilon}{2}}{n}\cdot\frac{1-\beta_2}{mA_{\max}}\frac{w_j}{\delta_j{}^\alpha}\xi_j(x^0).$$

Recalling that $\frac{w_j}{\delta_j{}^\alpha} = C \geq 2w_{\max}n^2 mA_{\max}$, it further follows that:

$$(x_p^0)^\alpha \sum_{l=1}^m A_{lp}y_l(x^0) \geq 2\left(1-\frac{\varepsilon}{2}\right)(1-\beta_2)\cdot n \cdot w_{\max}\cdot\xi_j(x^0). \tag{19}$$

Because $\varepsilon \leq \frac{1}{6}$ and $\beta_2 < \beta = \frac{\gamma}{5(\kappa+1)} = \frac{\varepsilon}{20(\kappa+1)} < \frac{\varepsilon}{20}$, it follows that $2\left(1-\frac{\varepsilon}{2}\right)(1-\beta_2) > 1$. Therefore:

$$\frac{(x_p^0)^\alpha \sum_{l=1}^m A_{lp}y_l(x^0)}{w_p} \geq \frac{(x_p^0)^\alpha \sum_{l=1}^m A_{lp}y_l(x^0)}{w_{\max}} > n\cdot\xi_j(x^0) = n\cdot\frac{(x_j^0)^\alpha \sum_{l=1}^m A_{lj}y_i(x^0)}{w_j}. \tag{20}$$

As $\alpha < 1$, we have that $\delta_j{}^\alpha > \delta_j$, and $\frac{w_j}{\delta_j} > \frac{w_j}{\delta_j{}^\alpha} = C$. Similar to (16), we can lower-bound y_i as:

$$y_i(x) \geq \frac{1-\beta_2}{mA_{\max}}\cdot\frac{w_j}{\delta_j}\cdot\frac{x_j^0\sum_i y_i(x)A_{ij}}{w_j} > \frac{1-\beta_2}{mA_{\max}}\cdot\frac{w_j}{\delta_j{}^\alpha}\cdot\frac{x_j^0\sum_i y_i(x)A_{ij}}{w_j}. \tag{21}$$

Then, recalling $A_{ip}x_p^0 > \frac{1-\frac{\varepsilon}{2}}{n}$, and using (21), it is simple to show that:

$$x_p^0\sum_l y_l(x^0)A_{lp} > n\cdot x_j^0\sum_{l=1}^m A_{lj}y_l(x^0). \tag{22}$$

As $\xi_j(x^0) \geq (1+\gamma)$ and $x_p^0 > \frac{\delta_p}{1-\beta_2}$, it immediately follows from (20) that x_p decreases by a factor $(1-\beta_2)$.

In the rest of the proof we show that (20) and (22) imply that the increase in the potential due to the decrease of variable x_p dominates the increase in the potential due to the decrease of variable x_j by at least a factor n. This result then further implies that the increase in the potential due to the decrease of variable x_p dominates the increase in the potential due to the decrease of *all* small x_k's that appear in the constraint

14

i (x_k's are such that $A_{ik} \neq 0$, $x_k^0 < \frac{\delta_k}{1-\beta_2}$, and $\frac{(x_k^0)^\alpha \sum_l y_l(x) A_{lk}}{w_k} \geq 1 + \gamma$).

Consider the following two cases: $w_p(x_p^0)^{1-\alpha} \geq (w_j x_j^0)^{1-\alpha}$ and $w_p(x_p^0)^{1-\alpha} < (w_j x_j^0)^{1-\alpha}$.

Case 1: $w_p(x_p^0)^{1-\alpha} \geq (w_j x_j^0)^{1-\alpha}$. Then, using (20):

$$w_p(x_p^0)^{1-\alpha}\left(\frac{(x_p^0)^\alpha \sum_{l=1}^m A_{lp} y_l(x^0)}{w_p} - (1+\gamma)\right) \geq (w_j x_j^0)^{1-\alpha}\left(\frac{(x_p^0)^\alpha \sum_{l=1}^m A_{lp} y_l(x^0)}{w_p} - (1+\gamma)\right)$$

$$\geq (w_j x_j^0)^{1-\alpha}\left(n \cdot \frac{(x_j^0)^\alpha \sum_{l=1}^m A_{lj} y_l(x^0)}{w_j} - (1+\gamma)\right)$$

$$\geq n \cdot (w_j x_j^0)^{1-\alpha}\left(\frac{(x_j^0)^\alpha \sum_{l=1}^m A_{lj} y_l(x^0)}{w_j} - (1+\gamma)\right).$$

$$(23)$$

Case 2: $w_p(x_p^0)^{1-\alpha} < (w_j x_j^0)^{1-\alpha}$. Then, using (22):

$$w_p(x_p^0)^{1-\alpha}\left(\frac{(x_p^0)^\alpha \sum_{l=1}^m A_{lp} y_l(x^0)}{w_p} - (1+\gamma)\right) = x_p^0 \sum_{l=1}^m A_{lp} y_l(x^0) - (1+\gamma) w_p(x_p^0)^{1-\alpha}$$

$$\geq x_p^0 \sum_{l=1}^m A_{lp} y_l(x^0) - (1+\gamma) w_j(x_j^0)^{1-\alpha}$$

$$\geq n \cdot x_j^0 \sum_{l=1}^m A_{lj} y_l(x^0) - (1+\gamma) w_j(x_j^0)^{1-\alpha}$$

$$\geq n \cdot (w_j x_j^0)^{1-\alpha}\left(\frac{(x_j^0)^\alpha \sum_{l=1}^m A_{lj} y_l(x^0)}{w_j} - (1+\gamma)\right).$$

$$(24)$$

Combining (23) and (24) with (15), it follows that:

$$\Phi(x^1) - \Phi(x^0) \geq \Omega(\beta_2) \sum_{j \in S^-} w_j(x_j^0)^{1-\alpha}\left(\frac{(x_j^0)^\alpha \sum_{i=1}^m y_i(x^0) A_{ij}}{w_j} - (1+\gamma)\right).$$

Finally, since for $j \notin S^-$: $\left(\frac{(x_j^0)^\alpha \sum_{i=1}^m y_i(x^0) A_{ij}}{w_j} - (1+\gamma)\right) < 0$:

$$\Phi(x^1) - \Phi(x^0) \geq \Omega(\beta_2) \sum_{j=1}^n w_j(x_j^0)^{1-\alpha}\left(\frac{(x_j^0)^\alpha \sum_{i=1}^m y_i(x^0) A_{ij}}{w_j} - (1+\gamma)\right)$$

$$= \Omega\left(\frac{\beta^2}{\ln(1/\delta_{\min})}\right)\left(\sum_{j=1}^n x_j^0 \sum_{i=1}^m y_i(x^0) A_{ij} - (1+\gamma) \sum_{j=1}^n w_j(x_j^0)^{1-\alpha}\right),$$

completing the proof. $\qquad \square$

Parts 2 and 3 of Lemma [7] appear in a somewhat similar form in [7]. However, part 3 requires significant additional results for bounding the potential change due to decrease of small x_j's (i.e., x_j's that are smaller than $\frac{\delta_j}{1-\beta}$) that were not needed in [7]. The rest of the results in this paper are new.

Consider the following definition of a stationary round:

Definition 4.11. *(Stationary round.) Let $S^- = \{j : x_j$ decreases$\}$. A round is stationary if it happens after the initial $\tau_0 + \tau_1$ rounds, where $\tau_0 = \frac{1}{\beta}\ln(\frac{1}{\delta_{\min}})$ and $\tau_1 = \frac{1}{\beta_2}\ln(nA_{\max})$, and both of the following two conditions hold:*

   1. $\sum_{j \in S^-} w_j x_j^{1-\alpha} \leq \gamma \sum_{j=1}^n w_j x_j^{1-\alpha}$, and

15

2. $\sum_{j=1}^{n} x_j \sum_{i=1}^{m} y_i(x) A_{ij} \leq (1 + 5\gamma/4) \sum_{j=1}^{n} w_j x_j^{1-\alpha}$.

In the rest of the proof, we first show that in any stationary round, we have an $O(\varepsilon)-$approximate solution, while in any non-stationary round, the potential function increases substantially.

We first prove the following lemma, which we will then be used in bounding the duality gap.

Lemma 4.12. *After the initial $\tau_0 + \tau_1$ rounds, where $\tau_0 = \frac{1}{\beta} \ln(\frac{1}{\delta_{\min}})$ and $\tau_1 = \frac{1}{\beta_2} \ln(n A_{\max})$, in each round of the algorithm: $\xi_j(x) \equiv \frac{x_j{}^{\alpha} \sum_i y_i(x) A_{ij}}{w_j} > 1 - \frac{5\gamma}{4}, \forall j$.*

Proof. Suppose without loss of generality that the algorithm starts with a feasible solution. This assumption is w.l.o.g. because, from Lemma 4.5, after at most τ_1 rounds the algorithm reaches a feasible solution, and from Lemma 4.4, once the algorithm reaches a feasible solution, it always maintains a feasible solution.

Choose any j. Using the same argument as in the proof of Lemma 4.4, after at most $\frac{1}{\beta} \ln(\frac{1}{\delta_j}) \leq \tau_0$ rounds, there exists at least one round in which $\xi_j(x) > 1 - \gamma$ (otherwise $x_j > 1$, which is a contradiction).

Observe that in any round for which $\xi_j(x) \leq 1 - \gamma$, x_j increases by a factor $1 + \beta_1 = 1 + \beta$. Therefore, the maximum number of consecutive rounds in which $\xi_j(x) \leq 1 - \gamma$ is at most $\frac{1}{\beta} \ln(\frac{1}{\delta_j}) \leq \tau_0$, otherwise x_j would increase to a value larger than 1, making x infeasible, which is a contradiction due to Lemma 4.4. The maximum amount by which $\xi_j(x)$ can decrease in any round is bounded by a factor $1 - \frac{\gamma}{4} \cdot \frac{\beta}{\ln(1/\delta_{\min})} = 1 - \frac{\gamma}{4} \cdot \frac{1}{\tau_0}$. Therefore, using the generalized Bernoulli's inequality, it follows that in any round:

$$\xi_j(x) \geq (1 - \gamma) \cdot \left(1 - \frac{\gamma}{4} \cdot \frac{1}{\tau_0}\right)^{\tau_0} \geq (1 - \gamma) \cdot \left(1 - \frac{\gamma}{4}\right) > 1 - \frac{5\gamma}{4}.$$

$\square$

A simple corollary of Lemma 4.12 is that:

Corollary 4.13. *After the initial $\tau_0 + \tau_1$ rounds, where $\tau_0 = \frac{1}{\beta} \ln(\frac{1}{\delta_{\min}})$ and $\tau_1 = \frac{1}{\beta_2} \ln(n A_{\max})$, in each round of the algorithm: $\sum_j x_j \sum_i y_i(x) A_{ij} > \left(1 - \frac{5\gamma}{4}\right) \sum_j w_j x_j^{1-\alpha}$.*

Proof. From Lemma 4.12, after the initial $\tau_0 + \tau_1$ rounds, it always holds $\xi_j(x) \equiv \frac{x_j{}^{\alpha} \sum_i y_i(x) A_{ij}}{w_j} \geq 1 - \frac{5\gamma}{4}$, $\forall j$. Multiplying both sides of the inequality by $w_j x_j^{1-\alpha}$, $\forall j$ and summing over j, the result follows. $\square$

Recall that $p_\alpha(x) \equiv \sum_j w_j f_\alpha(x_j)$ denotes the primal objective. The following lemma states that any stationary round holds an $(1 + 6\varepsilon)$-approximate solution.

Lemma 4.14. *In any stationary round: $p(x^*) \leq (1 + 6\varepsilon)p(x)$, where x^* is the optimal solution to (P_α).*

Proof. Since, by definition, a stationary round can only happen after the initial $\tau_0 + \tau_1$ rounds, we have that x in that round is feasible, and also from Lemma 4.6: $\sum_i y_i \leq (1 + 3\varepsilon) \sum_j x_j \sum_i y_i(x) A_{ij}$. Therefore, recalling Eq. (2) for the duality gap and denoting $\xi_j(x) = \frac{x_j{}^{\alpha} \sum_i y_i(x) A_{ij}}{w_j}$, we have that:

$$p(x^*) - p(x) \leq G(x, y(x)) = \sum_j w_j \frac{x_j^{1-\alpha}}{1-\alpha} \left(\xi_j^{-\frac{1-\alpha}{\alpha}} - 1\right) + \sum_i y_i(x) - \sum_j w_j x_j^{1-\alpha} \xi_j^{-\frac{1-\alpha}{\alpha}}$$

$$= \sum_j w_j \frac{x_j^{1-\alpha}}{1-\alpha} \left(\alpha \xi_j^{-\frac{1-\alpha}{\alpha}} - 1\right) + \sum_i y_i(x)$$

$$\leq \sum_j w_j \frac{x_j^{1-\alpha}}{1-\alpha} \left(\alpha \xi_j^{-\frac{1-\alpha}{\alpha}} - 1\right) + (1 + 3\varepsilon) \sum_j x_j \sum_i y_i(x) A_{ij}. \quad (25)$$

From Lemma 4.12, $\xi_j > 1 - \frac{5\gamma}{4}$, $\forall j$. Partition the indices of all the variables as follows:

$$S_1 = \left\{j : \xi_j \in \left(1 - \frac{5\gamma}{4}, 1 + \frac{5\gamma}{4}\right)\right\}, \quad S_2 = \left\{j : \xi_j \geq 1 + \frac{5\gamma}{4}\right\}.$$

Then, using (25):

$$p(x^*) - p(x) \leq G_1(x) + G_2(x),$$

16

where:

$$G_1(x) = \sum_{j \in S_1} w_j \frac{x_j^{1-\alpha}}{1-\alpha} \left(\alpha \xi_j^{-\frac{1-\alpha}{\alpha}} - 1 \right) + (1 + 3\varepsilon) \sum_{j \in S_1} x_j \sum_i y_i(x) A_{ij}$$

and

$$G_2(x) = \sum_{j \in S_2} w_j \frac{x_j^{1-\alpha}}{1-\alpha} \left(\alpha \xi_j^{-\frac{1-\alpha}{\alpha}} - 1 \right) + (1 + 3\varepsilon) \sum_{j \in S_2} x_j \sum_i y_i(x) A_{ij}.$$

The rest of the proof follows by upper-bounding $G_1(x)$ and $G_2(x)$.

Bounding $G_1(x)$. Observing that $\forall j$: $x_j \sum_i y_i(x) A_{ij} = w_j x_j^{1-\alpha} \xi_j$, we can write $G_1(x)$ as:

$$G_1(x) = \sum_{j \in S_1} w_j \frac{x_j^{1-\alpha}}{1-\alpha} \left(\alpha \xi_j^{-\frac{1-\alpha}{\alpha}} + (1 + 3\varepsilon)(1-\alpha)\xi_j - 1 \right). \tag{26}$$

Denote $r(\xi_j) = \alpha \xi_j^{-\frac{1-\alpha}{\alpha}} + (1 + 3\varepsilon)(1-\alpha)\xi_j - 1$. It is simple to verify that $r(\xi_j)$ is a convex function. Since $\xi_j \in \left(1 - \frac{5\gamma}{4}, 1 + \frac{5\gamma}{4} \right), \forall j \in S_1$, it follows that $r(\xi_j) < \max\{r(1 - 5\gamma/4), r(1 + 5\gamma/4)\}$. Now:

$$r(1 - 5\gamma/4) = \alpha \left(1 - \frac{5\gamma}{4} \right)^{-\frac{1-\alpha}{\alpha}} + (1-\alpha)(1 + 3\varepsilon)\left(1 - \frac{5\gamma}{4} \right) - 1$$

$$< \alpha \left(1 - \frac{5\gamma}{4} \right)^{-\frac{1-\alpha}{\alpha}} + (1-\alpha)(1 + 3\varepsilon) - 1.$$

If $\frac{1-\alpha}{\alpha} \le 1$, then as $(1 - 5\gamma/4)^{-1} \le (1 + 2\gamma)$, it follows that $(1 - 5\gamma/4)^{-\frac{1-\alpha}{\alpha}} \le 1 + 2\gamma$. Therefore:

$$r(1 - 5\gamma/4) < \alpha(1 + 2\gamma) + (1-\alpha)(1 + 3\varepsilon) - 1$$

$$= 2\gamma\alpha + 3 \cdot (1-\alpha)\varepsilon = \alpha\frac{\varepsilon}{2} + 3 \cdot (1-\alpha)\varepsilon$$

$$= 3\varepsilon \left(1 - \frac{5}{6}\alpha \right). \tag{27}$$

If $\frac{1-\alpha}{\alpha} > 1$, then (using generalized Bernoulli's inequality and $\varepsilon \le \frac{\alpha}{1-\alpha}$):

$$r(1 - 5\gamma/4) < \alpha\frac{1}{(1 - 5\gamma/4)^{\frac{1-\alpha}{\alpha}}} + (1-\alpha)(1 + 3\varepsilon) - 1$$

$$\le \alpha\frac{1}{1 - \frac{5\gamma}{4} \cdot \frac{1-\alpha}{\alpha}} + (1-\alpha)(1 + 3\varepsilon) - 1$$

$$\le \alpha\left(1 + \frac{5\gamma}{4} \cdot \frac{1-\alpha}{\alpha} \right) + (1-\alpha)(1 + 3\varepsilon) - 1$$

$$\le (1-\alpha)\left(\frac{5\gamma}{4} + 3\varepsilon \right)$$

$$< 4\varepsilon(1-\alpha). \tag{28}$$

On the other hand:

$$r(1 + 5\gamma) = \alpha\left(1 + \frac{5\gamma}{4} \right)^{-\frac{1-\alpha}{\alpha}} + (1-\alpha)(1 + 3\varepsilon)\left(1 + \frac{5\gamma}{4} \right) - 1$$

$$< \alpha + (1-\alpha)(1 + 4\varepsilon) - 1$$

$$= 4\varepsilon(1-\alpha). \tag{29}$$

Combining (27)–(29) with (26):

$$G_1(x) < 4\varepsilon \cdot \sum_{j \in S_1} w_j \frac{x_j^{1-\alpha}}{1-\alpha}. \tag{30}$$

Bounding $G_2(x)$. Because the round is stationary and $S_2 \subseteq S^-$, we have that: $\sum_{j \in S_2} w_j x_j^{1-\alpha} \le$

$\gamma \sum_{j=1}^{n} w_j x_j^{1-\alpha}$. Using the second part of the stationary round definition and that $\sum_{j \in S_2} x_j \sum_{i=1}^{m} y_i(x) A_{ij} > (1 - 5\gamma/4) \sum_{j \in S_2} w_j x_j^{1-\alpha}$ (follows from Lemma 4.12):

$$
\begin{aligned}
\sum_{j \in S_2} x_j \sum_{i=1}^{m} y_i(x) A_{ij} &= \sum_{k=1}^{m} x_k \sum_{i=1}^{m} y_i(x) A_{ik} - \sum_{l \notin S_2} x_l \sum_{l=1}^{m} y_l(x) A_{lk} \\
&\leq (1 + 5\gamma/4) \sum_{k=1}^{n} w_k x_k^{1-\alpha} - (1 - 5\gamma/4) \sum_{l \notin S_2} w_l x_l^{1-\alpha} \\
&\leq (1 + 5\gamma/4) \sum_{j \in S_2} w_j x_j^{1-\alpha} + \frac{5\gamma}{2} \sum_{l \notin S_2} w_l x_l^{1-\alpha} \\
&\leq \gamma(1 + 5\gamma/4) \sum_{k=1}^{n} w_k x_k^{1-\alpha} + \frac{5\gamma}{2} \sum_{k=1}^{n} w_k x_k^{1-\alpha} \\
&< 4\gamma \sum_{k=1}^{n} w_k x_k^{1-\alpha} = \varepsilon \sum_{k=1}^{n} w_k x_k^{1-\alpha}.
\end{aligned}
\tag{31}
$$

Above, first inequality follows from $\sum_{k=1}^{m} x_k \sum_{i=1}^{m} y_i(x) A_{ik} \leq (1 + 5\gamma/4) \sum_{k=1}^{n} w_k x_k^{1-\alpha}$ (part 2 of the stationary round definition) and Corollary 4.13. Second inequality follows by breaking the left summation into two summations: those with $j \in S_2$ and those with $l \notin S_2$. The third inequality follows from $S_2 \subseteq S$ and part 1 of the stationary round definition.

Observe that as $\xi_j \geq 1 + 5\gamma/4 > 1$, we have that $\xi_j^{-\frac{1-\alpha}{\alpha}} < 1$. Using (31), it follows that:

$$
\begin{aligned}
G_2(x) &= \sum_{j \in S_2} w_j \frac{x_j^{1-\alpha}}{1-\alpha} \left(\alpha \xi_j^{-\frac{1-\alpha}{\alpha}} - 1 \right) + (1 + 3\varepsilon) \sum_{j \in S_2} x_j \sum_i y_i(x) A_{ij} \\
&< (\alpha - 1) \sum_{j \in S_2} w_j \frac{x_j^{1-\alpha}}{1-\alpha} + (1 + 3\varepsilon) \sum_{j \in S_2} x_j \sum_i y_i(x) A_{ij} \\
&\leq (\alpha - 1) \sum_{j \in S_2} w_j \frac{x_j^{1-\alpha}}{1-\alpha} + \varepsilon(1 + 3\varepsilon) \sum_{k=1}^{n} w_k x_k^{1-\alpha} \\
&\leq -(1 - \alpha) \sum_{j \in S_2} w_j \frac{x_j^{1-\alpha}}{1-\alpha} + \frac{3}{2}\varepsilon(1 - \alpha) \sum_{k=1}^{n} w_k \frac{x_k^{1-\alpha}}{1-\alpha} \\
&< \frac{3}{2}\varepsilon(1 - \alpha) \sum_{k=1}^{n} w_k \frac{x_k^{1-\alpha}}{1-\alpha} \\
&< 2\varepsilon \sum_{k=1}^{n} w_k \frac{x_k^{1-\alpha}}{1-\alpha}.
\end{aligned}
\tag{32}
$$

Finally, combining (30) and (32):

$$
p(x^*) - p(x) < \left(4\varepsilon + 2\varepsilon \right) \sum_{j \in S_1} w_j \frac{x_j^{1-\alpha}}{1-\alpha}
$$

$$
= 6\varepsilon p(x).
$$

$\square$

Proof of Theorem 4.1. From Lemma 4.14, in any stationary round: $p(x^*) \leq p(x)(1 + 6\varepsilon)$. Therefore, to prove the theorem, it suffices to show that there are at most $O\left(\frac{1}{\alpha^2 \varepsilon^5} \ln^2 \left(R_w mn A_{\max} \right) \ln^2 \left(R_w \frac{mn A_{\max}}{\varepsilon} \right) \right)$ non-stationary rounds in total, where $R_w = w_{\max}/w_{\min}$, because we can always run the algorithm for

$\varepsilon' = \varepsilon/6$ to get an $\varepsilon-$approximation, and this would only affect the constant in the convergence time.

To bound the number of non-stationary rounds, we will show that the potential increases by a "large enough" multiplicative value in all the non-stationary rounds in which the potential is not too "small". For the non-stationary rounds in which the value of the potential is "small", we show that the potential increases by a large enough value so that there can be only few such rounds.

In the rest of the proof, we assume that the initial $\tau_0 + \tau_1$ rounds have passed, so that x is feasible, and the statement of Lemma 4.6 holds. This does not affect the overall bound on the convergence time, as

$$\tau_0 + \tau_1 = \frac{1}{\beta} \ln \left(\frac{1}{\delta_{\min}} \right) + \frac{1}{\beta_2} \ln(nA_{max}) = O \left(\frac{1}{\beta^2} \ln(nA_{\max}) \ln \left(\frac{1}{\delta_{\min}} \right) \right)$$

$$= O \left(\frac{1}{\alpha \varepsilon^4} \ln(nA_{\max}) \ln^2 \left(R_w \frac{mnA_{\max}}{\varepsilon} \right) \ln \left(R_w mnA_{\max} \right) \right). \tag{33}$$

To bound the minimum and the maximum values of the potential Φ, we will bound $\sum_j w_j \frac{x_j^{1-\alpha}}{1-\alpha}$ and $\frac{1}{\kappa} \sum_i y_i(x)$. Recall that $\Phi(x) = \sum_j w_j \frac{x_j^{1-\alpha}}{1-\alpha} - \frac{1}{\kappa} \sum_i y_i(x)$.

Since $\delta_j = \left(\frac{w_j}{2w_{\max}n^2 mA_{\max}} \right)^{\frac{1}{\alpha}} \geq \left(\frac{w_{\min}}{2w_{\max}n^2 mA_{\max}} \right)^{\frac{1}{\alpha}}$, x is always feasible, and $x_j \leq 1, \forall j$, we have that:

$$\frac{W}{1-\alpha} \cdot \left(\frac{w_{\min}}{2w_{\max}n^2 mA_{\max}} \right)^{\frac{1-\alpha}{\alpha}} \leq \sum_j w_j \frac{x_j^{1-\alpha}}{1-\alpha} \leq \frac{W}{1-\alpha}, \tag{34}$$

and

$$0 < \frac{1}{\kappa} \sum_i y_i(x) \leq \frac{Cm}{\kappa}. \tag{35}$$

Thus, we have:

$$\Phi_{\min} \geq -\frac{1}{\kappa} \sum_i y_i(x)$$

$$\geq -\frac{1}{\kappa} \cdot m \cdot C$$

$$\geq -O(m^2 n^2 A_{\max} w_{\max}), \tag{36}$$

and

$$\Phi_{\max} \leq \sum_{j=1}^n w_j \frac{1}{1-\alpha} = \frac{W}{1-\alpha}. \tag{37}$$

Recall from Lemma 4.9 that the potential never decreases. We consider the following three cases for the value of the potential:

Case 1: $\Phi_{\min} \leq \Phi \leq -\Theta(\frac{w_{\min}}{A_{\max}})$. Since in this case $\Phi < 0$, we have that $\sum_i y_i(x) > \kappa \sum_j w_j \frac{x_j^{1-\alpha}}{1-\alpha}$. From Lemma 4.6, $\sum_j x_j \sum_i y_i(x) A_{ij} \geq (1 - 3\varepsilon) \sum_j w_j \frac{x_j^{1-\alpha}}{1-\alpha}$, thus implying:

$$\sum_j x_j \sum_i y_i(x) A_{ij} \geq \frac{1 - 3\varepsilon}{\kappa} \sum_j w_j \frac{x_j^{1-\alpha}}{1-\alpha} \geq 2 \cdot \sum_j w_j \frac{x_j^{1-\alpha}}{1-\alpha}, \tag{38}$$

as $\kappa \geq \frac{1}{\varepsilon}$ and $\varepsilon \leq \frac{1}{6}$. Combining Part 3 of Lemma 4.10 and (38), the potential increases by at least:

$$\Omega \left(\frac{\beta^2}{\ln(1/\delta_{\min})} \right) \sum_j x_j \sum_i y_i(x) A_{ij} = \left(\frac{\beta^2}{\ln(1/\delta_{\min})} \right) \sum_i y_i(x) = \left(\frac{\beta^2}{\ln(1/\delta_{\min})} \cdot \kappa \right) (-\Phi(x))$$

$$= \Omega \left(\frac{\gamma^2}{\kappa \ln(1/\delta_{\min})} \right) (-\Phi(x)).$$

19

Since the potential never decreases, there can be at most
$$O\left(\frac{\kappa \ln(1/\delta_{\min})}{\gamma^2} \ln\left(\frac{-\Phi_{\min}}{w_{\min}/A_{\max}}\right)\right) = O\left(\frac{1}{\alpha}\frac{1}{\varepsilon^3} \ln^2\left(R_w nmA_{\max}\right) \ln\left(R_w \frac{nmA_{\max}}{\varepsilon}\right)\right)$$
Case 1 rounds.

Case 2: $-O\left(\frac{w_{\min}}{A_{\max}}\right) < \Phi \leq O\left(\frac{W}{1-\alpha} \cdot \left(\frac{w_{\min}}{2w_{\max}n^2mA_{\max}}\right)^{\frac{1-\alpha}{\alpha}}\right)$. From Lemma 4.6, there exists at least one i such that $\sum_j A_{ij}x_j \geq 1 - (1+1/\kappa)\varepsilon$. Since $A_{ij} \leq A_{\max} \ \forall i,j$, it is also true that $\sum_j x_j \geq \frac{1-(1+1/\kappa)\varepsilon}{A_{\max}}$, and as $x_j^{1-\alpha} \geq x_j$ and $\kappa \geq \frac{1}{\varepsilon}$, it follows that $\sum_j w_j x_j^{1-\alpha} \geq (1 - \varepsilon(1+\varepsilon))\left(\frac{w_{\min}}{A_{\max}}\right)$. From (34), we also have $\sum_j w_j \frac{x_j^{1-\alpha}}{1-\alpha} \geq \frac{W}{1-\alpha} \cdot \left(\frac{w_{\min}}{2w_{\max}n^2mA_{\max}}\right)^{\frac{1-\alpha}{\alpha}}$. Therefore:

$$\sum_j w_j \frac{x_j^{1-\alpha}}{1-\alpha} \geq \max\left\{(1-\varepsilon(1+\varepsilon))\frac{1}{1-\alpha} \cdot \frac{w_{\min}}{A_{\max}}, \ \frac{W}{1-\alpha} \cdot \left(\frac{w_{\min}}{2w_{\max}n^2mA_{\max}}\right)^{\frac{1-\alpha}{\alpha}}\right\}. \tag{39}$$

If $\Phi \leq \frac{1}{10} \cdot \max\left\{(1-\varepsilon(1+\varepsilon))\frac{1}{1-\alpha} \cdot \frac{w_{\min}}{A_{\max}}, \ \frac{W}{1-\alpha} \cdot \left(\frac{w_{\min}}{2w_{\max}n^2mA_{\max}}\right)^{\frac{1-\alpha}{\alpha}}\right\}$, then

$$\sum_i y_i(x) \geq \frac{9}{10}\kappa \cdot \frac{1}{1-\alpha} \sum_j w_j x_j^{\alpha}$$

$$\geq \frac{9}{10}\kappa \cdot \max\left\{(1-\varepsilon(1+\varepsilon))\frac{1}{1-\alpha}\frac{w_{\min}}{A_{\max}}, \ \frac{W}{1-\alpha} \cdot \left(\frac{w_{\min}}{2w_{\max}n^2mA_{\max}}\right)^{\frac{1-\alpha}{\alpha}}\right\}.$$

From Lemma 4.6,
$$\sum_i y_i(x) \sum_j A_{ij}x_j \geq (1-3\varepsilon)\sum_i y_i(x)$$

$$\geq (1-3\varepsilon)\frac{9}{10}\kappa \cdot \max\left\{(1-\varepsilon(1+\varepsilon))\frac{1}{1-\alpha} \cdot \frac{w_{\min}}{A_{\max}}, \ \frac{W}{1-\alpha} \cdot \left(\frac{w_{\min}}{2w_{\max}n^2mA_{\max}}\right)^{\frac{1-\alpha}{\alpha}}\right\}.$$

From the third part of Lemma 4.10, the potential increases additively by at least
$$\Omega\left(\frac{\beta^2\kappa}{\ln(1/\delta_{\min})}\right) \cdot \max\left\{\frac{1}{1-\alpha} \cdot \frac{w_{\min}}{A_{\max}}, \ \frac{W}{1-\alpha} \cdot \left(\frac{w_{\min}}{2w_{\max}n^2mA_{\max}}\right)^{\frac{1-\alpha}{\alpha}}\right\},$$

and, therefore, $\Phi = \Omega\left(\frac{W}{1-\alpha} \cdot \left(\frac{w_{\min}}{2w_{\max}n^2mA_{\max}}\right)^{\frac{1-\alpha}{\alpha}}\right)$ after at most
$$O\left(\frac{\ln(1/\delta_{\min})\kappa}{\gamma^2}\right) = O\left(\frac{1}{\alpha}\frac{1}{\varepsilon^3} \ln\left(R_w nmA_{\max}\right) \ln\left(R_w \frac{nmA_{\max}}{\varepsilon}\right)\right)$$
rounds.

Case 3: $\Omega\left(\frac{W}{1-\alpha} \cdot \left(\frac{w_{\min}}{2w_{\max}n^2mA_{\max}}\right)^{\frac{1-\alpha}{\alpha}}\right) \leq \Phi \leq \frac{W}{1-\alpha}$. In this case, $\Phi = O\left(\sum_j w_j \frac{x_j^{1-\alpha}}{1-\alpha}\right)$. If the round is stationary, then from Lemma 4.14, $p(x^*) \leq (1+6\varepsilon)p(x)$. If the round is not stationary, then from Definition 4.11, either:

1. $\sum_{k\in S^-} w_k x_k^{1-\alpha} > \gamma \sum_{j=1}^n w_j x_j^{1-\alpha}$, or

2. $\sum_{j=1}^n x_j \sum_{i=1}^m y_i(x) A_{ij} > (1 + \frac{5\gamma}{4}) \sum_{j=1}^n w_j x_j^{1-\alpha}$.

If the former is true, then using the first part of Lemma 4.10, the potential increases by at least $\Omega\left(\frac{\beta^2\gamma}{\ln(1/\delta_{\min})}\right) \cdot \sum_j w_j x_j^{1-\alpha} = \Omega\left(\frac{\beta^2\gamma}{\ln(1/\delta_{\min})}\right) \cdot (1-\alpha)\Phi$. If the latter is true, from the third part of Lemma 4.10, the potential increases by at least $\Omega\left(\frac{\beta^2\gamma}{\ln(1/\delta_{\min})}\right) \cdot \sum_j w_j x_j^{1-\alpha} = \Omega\left(\frac{\beta^2\gamma}{\ln(1/\delta_{\min})}\right) \cdot (1-\alpha)\Phi$. It follows that there are at

most

$$O\left(\frac{1}{1-\alpha} \cdot \frac{\ln(1/\delta_{\min})}{\beta^2 \gamma} \ln\left(\frac{\frac{W}{1-\alpha}}{\frac{W}{1-\alpha} \cdot \left(\frac{w_{\min}}{2w_{\max} n^2 m A_{\max}}\right)^{\frac{1-\alpha}{\alpha}}}\right)\right)$$

$$= O\left(\frac{1}{\alpha^2} \frac{1}{\varepsilon^5} \ln^2\left(R_w \cdot mn A_{\max}\right) \ln^2\left(R_w \cdot \frac{mn A_{\max}}{\varepsilon}\right)\right)$$

non-stationary Case 3 rounds.

Combining the three cases with the bound on $\tau_0 + \tau_1$ (33), the total convergence time is at most:

$$O\left(\frac{1}{\alpha^2 \varepsilon^5} \ln^2\left(R_w mn A_{\max}\right) \ln^2\left(R_w \cdot \frac{mn A_{\max}}{\varepsilon}\right)\right)$$

rounds, as claimed. $\qquad\square$

4.2 Proof of Theorem 4.2

The proof outline for the convergence of α-FAIRPSOLVER in the $\alpha = 1$ case is as follows. First, we show that in any round it cannot be the case that only "small" x_j's (i.e., x_j's that are smaller than $\frac{\delta_j}{1-\beta}$) decrease. In fact, we show that the increase in the potential due to updates of "small" variables is dominated by the increase in the potential due to those variables that decrease multiplicatively by a factor $(1 - \beta_2) = (1 - \beta)$ (Lemmas 4.15 and 4.16). We then define a stationary round and show that: (i) in any non-stationary round the potential increases significantly, and (ii) in any stationary round, the solution x at the beginning of the round provides an additive $5W\varepsilon$–approximation to the optimum objective value.

Lemma 4.15. *Starting with a feasible solution, in any round of the algorithm:*

1. $\sum_{\{k \in S^- : x_k \geq \frac{\delta_k}{1-\beta}\}} x_k \sum_{i=1}^m y_i(x) A_{ik} \geq \frac{1}{2} \sum_{j \in S^-} x_j \sum_{i=1}^m y_i(x) A_{ij}$.

2. $\sum_{\{k \in S^- : x_k \geq \frac{\delta_k}{1-\beta}\}} \frac{x_k \sum_{i=1}^m y_i(x) A_{ik}}{w_k} \geq \frac{1}{2} \sum_{j \in S^-} \frac{x_j \sum_{i=1}^m y_i(x) A_{ij}}{w_j}$.

Proof. Fix any round, and let $x^0, y(x^0)$ and $x^1, y(x^1)$ denote the values of x, y at the beginning and at the end of the round, respectively. If for all $j \in S^-$ $x_j^0 \geq \frac{\delta_j}{1-\beta}$, there is nothing to prove.

Suppose that there exists some $x_j^0 < \frac{\delta_j}{1-\beta}$ that decreases. Then from Lemma 4.7 there exists at least one $i \in \{1, ..., m\}$ such that $A_{ij} \neq 0$, and:

- $\sum_{k=1}^n A_{ik} x_k^0 > 1 - \frac{\varepsilon}{2}$, and

- $y_i(x) \geq \frac{\sum_{l=1}^m y_l(x^0) A_{lj}}{m A_{\max}} > (1 - \beta) \frac{w_j}{\delta_j} \frac{1}{m A_{\max}} \frac{x_j^0 \sum_{l=1}^m y_l(x^0) A_{lj}}{w_j}$.

Since $\sum_{k=1}^n A_{ik} x_k^0 > 1 - \frac{\varepsilon}{2}$, there exists at least one p such that $A_{ip} x_p^0 > \frac{1-\frac{\varepsilon}{2}}{n}$. Recalling that $C = \frac{w_j}{\delta_j} \geq 2w_{\max} n^2 m A_{\max}$:

$$\begin{aligned}
(x_p^0) A_{ip} y_i(x^0) &> C \cdot \frac{(1-\beta)}{m A_{\max}} \cdot \frac{1-\frac{\varepsilon}{2}}{n} \cdot \frac{x_j^0 \sum_{l=1}^m y_l(x^0) A_{lj}}{w_j} \\
&> 2w_{\max} n^2 m A_{\max} \cdot \frac{(1-\beta)}{m A_{\max}} \cdot \frac{1-\frac{\varepsilon}{2}}{n} \cdot \frac{x_j^0 \sum_{l=1}^m y_l(x^0) A_{lj}}{w_j} \\
&\geq 2n w_{\max}(1-\beta)\left(1 - \frac{\varepsilon}{2}\right) \cdot \frac{x_j^0 \sum_{l=1}^m y_l(x^0) A_{lj}}{w_j} \\
&\geq n w_{\max} \frac{x_j^0 \sum_{l=1}^m y_l(x^0) A_{lj}}{w_j}. \tag{40}
\end{aligned}$$

21

Since x_j decreases, it must be $\frac{x_j^0 \sum_{l=1}^m y_l(x^0) A_{lj}}{w_j} \geq 1 + \gamma$. Using (40):

$$\frac{x_p^0 \sum_{l=1}^m y_l(x^0) A_{lp}}{w_p} \geq \frac{(x_p^0) A_{ip} y_i(x^0)}{w_{\max}} \geq n \frac{x_j^0 \sum_{l=1}^m y_l(x^0) A_{lj}}{w_j} \geq 1 + \gamma,$$

and, therefore, x_p decreases as well. Moreover, since (40) implies

$$x_p^0 \sum_{l=1}^m y_l(x^0) A_{lp} \geq \sum_{\{j \in S^- : x_j < \frac{\delta_j}{1-\beta} \wedge A_{ij} \neq 0\}} \frac{w_{\max}}{w_j} x_j^0 \sum_{l=1}^m y_l(x^0) A_{lj} \geq \sum_{\{j \in S^- : x_j < \frac{\delta_j}{1-\beta} \wedge A_{ij} \neq 0\}} x_j^0 \sum_{l=1}^m y_l(x^0) A_{lj},$$

the proof of the first part of the lemma follows. The second part follows from (40) as well, since:

$$\frac{x_p^0 \sum_{l=1}^m y_l(x^0) A_{lp}}{w_p} \geq \frac{(x_p^0) A_{ip} y_i(x^0)}{w_{\max}}$$
$$\geq n \frac{x_j^0 \sum_{l=1}^m y_l(x^0) A_{lj}}{w_j},$$

which, given that x_j was chosen arbitrarily, implies:

$$\frac{x_p^0 \sum_{l=1}^m y_l(x^0) A_{lp}}{w_p} \geq \sum_{\{j \in S^- : x_k < \frac{\delta_k}{1-\beta} \wedge A_{ik} \neq 0\}} \frac{x_k^0 \sum_{l=1}^m y_l(x^0) A_{lj}}{w_k}.$$

$\square$

Lemma 4.16. *Let $x^0, y(x^0)$ and $x^1, y(x^1)$ denote the values of x, y at the beginning and at the end of any fixed round, respectively. If x^0 is feasible, then the potential increase in the round is at least:*

1. $\Phi(x^1) - \Phi(x^0) \geq \Omega(\beta\gamma) \sum_{j \in S^+} w_j$;

2. $\Phi(x^1) - \Phi(x^0) \geq \Omega(\beta) \left((1 - \gamma)W - \sum_{j=1}^n x_j^0 \sum_{i=1}^m y_i(x^0) A_{ij} \right)$.

3. $\Phi(x^1) - \Phi(x^0) \geq \Omega(\beta) \left(\sum_{j=1}^n x_j^0 \sum_{i=1}^m y_i(x^0) A_{ij} - (1 + \gamma)W \right)$.

Proof.
Proof of 1: Recall that:

$$\Phi(x^1) - \Phi(x^0) \geq \sum_{j=1}^n w_j \frac{x_j^1 - x_j^0}{x_j^1} \left(1 - \frac{x_j^1 \sum_{i=1}^m y_i(x^1) A_{ij}}{w_j} \right) \geq \sum_{j \in S^+} w_j \frac{x_j^1 - x_j^0}{x_j^1} \left(1 - \frac{x_j^1 \sum_{i=1}^m y_i(x^1) A_{ij}}{w_j} \right).$$

Let $\xi_j(x^1) = \frac{x_j^1 \sum_{i=1}^m y_i(x^1) A_{ij}}{w_j}$, $\xi_j(x^0) = \frac{x_j^0 \sum_{i=1}^m y_i(x^0) A_{ij}}{w_j}$.

If $j \in S^+$, then $x_j^1 = (1+\beta)x_j^0$ and $\xi_j(x^0) \leq 1 - \gamma$. Since from the choice of parameters ξ_j increases by at most a factor of $1 + \gamma/4$, it follows that: $\xi_j(x^1) \leq (1-\gamma)(1+\gamma/4) \leq 1 - \frac{3}{4}\gamma$, which gives $1 - \xi_j(x^1) \geq \frac{3}{4}\gamma$. Therefore:

$$\Phi(x^1) - \Phi(x^0) \geq \frac{\beta}{1+\beta} \cdot \frac{3}{4}\gamma \cdot \sum_{j \in S^+} w_j.$$

Proof of 2: The proof is equivalent to the proof of the second part of Lemma 4.10 and is omitted.
Proof of 3: Using that for $j \in S^-$ we have that $\frac{x_j^0 \sum_{i=1}^m y_i(x^0) A_{ij}}{w_j} \geq 1 + \gamma$ and $x_j^1 = \max\{(1-\beta)x_j^0, \delta_j\}$,

we can lower bound the increase in the potential as:

$$\Phi(x^1) - \Phi(x^0) \geq \sum_{\{j \in S^- : x_j^0 \geq \frac{\delta_j}{1-\beta}\}} w_j \frac{x_j^1 - x_j^0}{x_j^1}\left(1 - \frac{x_j^1 \sum_{i=1}^m y_i(x^1) A_{ij}}{w_j}\right)$$

$$= \frac{\beta}{1-\beta} \sum_{\{j \in S^- : x_j^0 \geq \frac{\delta_j}{1-\beta}\}} w_j\left(\frac{x_j^1 \sum_{i=1}^m y_i(x^1) A_{ij}}{w_j} - 1\right)$$

$$\geq \frac{\beta}{1-\beta} \sum_{\{j \in S^- : x_j^0 \geq \frac{\delta_j}{1-\beta}\}} w_j\left((1-\gamma/4)\frac{x_j^0 \sum_{i=1}^m y_i(x^0) A_{ij}}{w_j} - 1\right)$$

$$\geq \frac{\beta}{1-\beta}(1-\gamma/4) \sum_{\{j \in S^- : x_j^0 \geq \frac{\delta_j}{1-\beta}\}} w_j\left(\frac{x_j^0 \sum_{i=1}^m y_i(x^0) A_{ij}}{w_j} - (1+\gamma)\right). \quad (41)$$

Now consider $k \in S^-$ such that $x_k^0 < \frac{\delta_k}{1-\beta}$. From the proof of Lemma 4.15, for each such x_k there exists a constraint i and a variable $x_p \geq \frac{\delta_p}{1-\beta}$ with $p \in S^-$ such that $A_{ik} \neq 0$, $A_{ip \neq 0}$, $x_p^0 \sum_l y_l(x^0) A_{lp} \geq n \cdot x_k^0 \sum_l y_l(x^0) A_{lk}$, and $\frac{x_p^0 \sum_l y_l(x^0) A_{lp}}{w_p} \geq n \cdot \frac{x_k^0 \sum_l y_l(x^0) A_{kp}}{w_k}$. If $w_k \leq w_p$ then

$$w_p\left(\frac{x_p^0 \sum_l y_l(x^0) A_{lp}}{w_p} - (1+\gamma)\right) \geq w_k\left(n \cdot \frac{x_k^0 \sum_l y_l(x^0) A_{kp}}{w_k} - (1+\gamma)\right)$$

$$\geq n \cdot w_k\left(\frac{x_k^0 \sum_l y_l(x^0) A_{kp}}{w_k} - (1+\gamma)\right).$$

On the other hand, if $w_k > w_p$, then:

$$w_p\left(\frac{x_p^0 \sum_l y_l(x^0) A_{lp}}{w_p} - (1+\gamma)\right) = (x_p^0 \sum_l y_l(x^0) A_{lp} - (1+\gamma)w_p)$$

$$> n \cdot x_k^0 \sum_l y_l(x^0) A_{kp} - (1+\gamma)w_k$$

$$\geq n \cdot w_k\left(\frac{x_k^0 \sum_l y_l(x^0) A_{kp}}{w_k} - (1+\gamma)\right).$$

It follows from (41) that:

$$\Phi(x^1) - \Phi(x^0) \geq \frac{\beta}{1-\beta}\frac{1-\gamma/4}{2} \sum_{j \in S^-} w_j\left(\frac{x_j^0 \sum_{i=1}^m y_i(x^0) A_{ij}}{w_j} - (1+\gamma)\right).$$

Finally, since for $j \notin S^-$ we have that $\frac{x_j^0 \sum_{i=1}^m y_i(x^0) A_{ij}}{w_j} < 1+\gamma$:

$$\Phi(x^1) - \Phi(x^0) \geq \frac{\beta}{1-\beta}\frac{1-\gamma/4}{2} \sum_{j=1}^n w_j\left(\frac{x_j^0 \sum_{i=1}^m y_i(x^0) A_{ij}}{w_j} - (1+\gamma)\right)$$

$$= \Omega(\beta)\left(\sum_{j=1}^n x_j^0 \sum_{i=1}^m y_i(x^0) A_{ij} - (1+\gamma)\sum_{j=1}^n w_j\right).$$

□

Consider the following definition of a stationary round:

Definition 4.17. *A round is stationary if it happens after the initial $\tau_0 + \tau_1$ rounds, where $\tau_0 = \frac{1}{\beta}\ln(1/\delta_{\min})$, $\tau_1 = \frac{1}{\beta}\ln(nA_{\max})$ and if both of the following conditions hold:*

- $\sum_{j \in S^+} w_j \leq W/\tau_0$;

- $(1 - 2\gamma)W \leq \sum_{j=1}^n x_j \sum_{i=1}^m y(x)A_{ij} \leq (1 + 2\gamma)W$.

We first show that in any non-stationary round there is a sufficient progress towards the $\varepsilon-$approximate solution.

Lemma 4.18. *In any non-stationary round the potential function increases by at least $\Omega(\beta\gamma \cdot W/\tau_0)$.*

Proof. A round is non-stationary if either of the two conditions from Definition 4.17 does not hold. If the first condition does not hold, then from the first part of Lemma 4.16, the potential increases by $\Omega(\beta\gamma \cdot W/\tau_0)$. If the second condition does not hold, then from either the second or the third part of Lemma 4.16 the potential increases by at least $\Omega(\beta\gamma W) \geq \Omega(\beta\gamma \cdot W/\tau_0)$. □

Before proving that in every non-stationary round, the solution is $O(\varepsilon)-$approximate, we will need the following intermediary lemma.

Lemma 4.19. *Starting with a feasible solution and after at most $\tau_0 = \frac{1}{\beta} \ln \left(\frac{1}{\delta_{\min}} \right)$ rounds, in any round of the algorithm:*

$$\min_j \frac{x_j \sum_{i=1}^m y_i(x)A_{ij}}{w_j} \geq (1 - \gamma)^{\tau_0}.$$

Proof. First, we claim that after the algorithm reaches a feasible solution it takes at most $\tau_0 + 1$ additional rounds for each agent j to reach a round in which $\frac{x_j \sum_{i=1}^m y_i(x)A_{ij}}{w_j} > 1 - \gamma$. Suppose not, and pick any agent k for which in each of the $\tau_0 + 1$ rounds following the first round that holds a feasible solution: $\frac{x_k \sum_{i=1}^m y_i(x)A_{ik}}{w_k} \leq 1 - \gamma$. Then x_k increases in each of the rounds and after $\frac{1}{\beta} \ln(\frac{1}{\delta_k}) \leq \tau_0$ rounds we have $x_k \geq 1$. Therefore, after at most $\tau_0 + 1$ rounds the solution becomes infeasible, which is a contradiction (due to Lemma 4.4).

Now choose any x_j and observe $\xi_j = \frac{x_j \sum_{i=1}^m y_i(x)A_{ij}}{w_j}$ over the rounds that happen after the first $\tau_0 + 1$ rounds. The maximum number of consecutive rounds for which $\xi_j \leq 1 - \gamma$ is $\tau_j = \frac{1}{\beta} \ln(\frac{1}{\delta_j}) \leq \tau_0$, otherwise we would have $x_j > 1$, a contradiction. Since in any round, due to the choice of the algorithm parameters, ξ_j decreases by at most a factor of $1 - \gamma/4$, the minimum value that ξ_j can take is at least $(1 - \gamma)(1 - \gamma/4)^{\tau_j/2} > (1 - \gamma)^{\tau_0}$, thus completing the proof. □

Now we are ready to prove that a solution in a stationary round is $O(\varepsilon)-$approximate.

Lemma 4.20. *In any stationary round: $p_1(x^*) - p_1(x) \leq 5\varepsilon W$, where x^* is the optimal solution.*

Proof. Since, due to Definition 4.17, a stationary round can only happen after the initial $\tau_0 + \tau_1$ rounds, we have that in any stationary round the solution is feasible (Lemmas 4.4 and 4.5) and approximate complementary slackness (Lemma 4.6) holds.

Recall the expression for the duality gap:

$$G_1(x, y) = - \sum_{j=1}^n w_j \ln \left(\frac{x_j \sum_{i=1}^m y_i A_{ij}}{w_j} \right) + \sum_{i=1}^m y_i - W.$$

From the second part of Lemma 4.6:

$$\sum_{i=1}^m y_i \leq (1 + 3\varepsilon) \sum_{j=1}^n x_j \sum_{i=1}^m y_i A_{ij}.$$

Therefore:

$$G_1(x, y) \leq - \sum_{j=1}^n w_j \ln \left(\frac{x_j \sum_{i=1}^m y_i A_{ij}}{w_j} \right) + (1 + 3\varepsilon) \sum_{j=1}^n x_j \sum_{i=1}^m y_i A_{ij} - W.$$

24

Since the round is stationary, we have that $\sum_{j=1}^{n} x_j \sum_{i=1}^{m} y_i A_{ij} \leq (1 + 2\gamma)W$, which gives:

$$G_1(x, y) \leq - \sum_{j=1}^{n} w_j \ln \left(\frac{x_j \sum_{i=1}^{m} y_i A_{ij}}{w_j} \right) + 4\varepsilon W. \tag{42}$$

Let $\xi_j = \frac{x_j \sum_{i=1}^{m} y_i A_{ij}}{w_j}$. The remaining part of the proof is to bound $- \sum_{j=1}^{n} w_j \ln(\xi_j) \leq - \sum_{j : \xi_j < 1} w_j \ln(\xi_j)$. For $\xi_j \in (1 - \gamma, 1)$, we have that $-w_j \ln(\xi_j) \leq \gamma w_j$. To bound the remaining terms, we will use Lemma 4.19 and the bound of the sum of the weights w_j for which $\xi_j \in S^+$ (that is, $\xi_j \leq 1 - \gamma$). It follows that:

$$\begin{aligned}
- \sum_{j=1}^{n} w_j \ln(\xi_j) &\leq - \sum_{k : \xi_k \in (1-\gamma, 1)} w_k \ln(\xi_k) - \sum_{l \in S^+} w_l \ln(\xi_l) \\
&\leq \gamma \sum_{k : \xi_k \in (1-\gamma, 1)} w_k - \ln\left((1 - \gamma)^{\tau_0} \right) \cdot \sum_{l \in S^+} w_l \quad \text{(from Lemma 4.19)} \\
&\leq \gamma W + \tau_0 \gamma \cdot \frac{W}{\tau_0} \\
&= 2\gamma W \\
&= \frac{\varepsilon}{2} W.
\end{aligned} \tag{43}$$

Combining (42) and (43), and recalling that $p_1(x^*) - p_1(x) \leq G_1(x, y(x))$, the result follows. $\qquad\square$

Proof of Theorem 4.2. Consider the values of the potential in the rounds following the initial $\tau_0 + \tau_1$ rounds, where $\tau_0 = \frac{1}{\beta} \ln(1/\delta_{\min})$, $\tau_1 = \frac{1}{\beta} \ln(n A_{\max})$ (so that the solution x is feasible in each round and the approximate complementary slackness holds). Observe that $\tau_0 + \tau_1 = o\left(\frac{\ln^2(nm A_{\max} R_w) \ln^2\left(\frac{nm A_{\max}}{\varepsilon} R_w \right)}{\varepsilon^5} \right)$.

We start by bounding the minimum and the maximum values that the potential can take. Recall (from Lemma 4.9) that the potential never decreases.

Due to Lemma 4.4, $x_j \in [\delta_j, 1]$, $\forall j$, and therefore we can bound the two summations in the potential as:

$$\sum_{j} w_j \ln(x_j) \geq \sum_{j} w_j \ln(\delta_j) = -O\left(W \cdot \ln\left(\frac{w_{\max}}{w_{\min}} nm A_{\max} \right) \right), \tag{44}$$

$$\sum_{j} w_j \ln(x_j) \leq \sum_{j} w_j \ln(1) \leq 0, \tag{45}$$

$$-\frac{1}{\kappa} \sum_{i} y_i(x) \geq -\frac{mC}{\kappa} \cdot e^0 > -mC = -O(w_{\max} n^2 m^2 A_{\max}), \tag{46}$$

and

$$-\frac{1}{\kappa} \sum_{i} y_i(x) < -\frac{mC}{\kappa} \cdot e^{-\kappa} < 0. \tag{47}$$

From (44) and (46):

$$\Phi_{\min} \geq -O(w_{\max} n^2 m^2 A_{\max}). \tag{48}$$

On the other hand, from (45) and (47):

$$\Phi_{\max} < 0. \tag{49}$$

Consider the following two cases:

Case 1: $\frac{1}{\kappa} \sum_{i} y_i(x) \geq W \cdot \ln\left(e \cdot \frac{w_{\max}}{w_{\min}} nm A_{\max} \right)$. Then $\frac{1}{\kappa} \sum_{i} y_i(x) \leq -\Phi(x) \leq \frac{2}{\kappa} \sum_{i} y_i(x)$ and $\frac{1}{\kappa} \sum_{i} y_i(x) \geq W$. From the third part of Lemma 4.6, we have that $\sum_{j} x_j \sum_{i} y_i(x) A_{ij} \geq (1 - 3\varepsilon) \sum_{i} y_i(x) \geq$

$2W$. Thus using the Part 2 of Lemma 4.16, we get that the potential increases by

$$\Omega(\beta) \cdot \sum_j x_j \sum_i y_i(x) A_{ij} = \Omega\left(\beta \cdot \sum_i y_i(x)\right) = \Omega(\beta\kappa) \cdot (-\Phi(x)).$$

Finally, since $\beta\kappa = \Theta(\gamma)$, there can be at most $O\left(\frac{1}{\gamma} \ln\left(\frac{R_w n m A_{\max}}{W \ln(R_w n m A_{\max})}\right)\right)$ Case 1 rounds.

Case 2: $\frac{1}{\kappa}\sum_i y_i(x) < W \cdot \ln\left(e \cdot \frac{w_{\max}}{w_{\min}} n m A_{\max}\right)$. Then $-2W \cdot \ln\left(e \cdot \frac{w_{\max}}{w_{\min}} n m A_{\max}\right) < \Phi(x) < 0$. From Lemma 4.20, if a round is stationary, then $p(x^*) - p(x) \leq 5\varepsilon W$. If a round is non-stationary, from Lemma 4.18, the potential increases (additively) by at least $\Omega(\beta\gamma \cdot W/\tau_0)$. Therefore, the maximum number of non-stationary rounds is at most:

$$O\left(\frac{W \ln(n m A_{\max} w_{\max}/w_{\min})}{\beta\gamma W/\tau_0}\right) = O\left(\frac{1}{\beta^2\gamma} \cdot \ln^2(R_w n m A_{\max})\right)$$

$$= O\left(\frac{\ln^2(R_w n m A_{\max}) \ln^2\left(R_w \frac{n m A_{\max}}{\varepsilon}\right)}{\varepsilon^5}\right).$$

Combining the results for the Case 1 and Case 2, the theorem follows by invoking α-FAIRPSOLVER for the approximation parameter $\varepsilon' = \varepsilon/5$. $\qquad\square$

4.3 Proof of Theorem 4.3

The outline of the proof of Theorem 4.3 is as follows. First, we show that in any round of the algorithm the variables that decrease by a multiplicative factor $(1 - \beta_2)$ dominate the potential increase due to *all the variables* that decrease (Lemma 4.21). This result is then used in Lemma 4.22 to show the appropriate lower bound on the potential increase. Observe that for $\alpha > 1$ the objective function $p_\alpha(x)$, and, consequently, the potential function $\Phi(x)$ is negative for any feasible x. To yield a poly-logarithmic convergence time in R_w, m, n, and $A_{\max}$, the idea is to show that the negative potential $-\Phi(x)$ decreases by some multiplicative factor whenever x is not a "good" approximation to x^* – the optimal solution to (P_α). This idea, combined with the fact that the potential never decreases (and therefore $-\Phi(x)$ never increases) and with upper and lower bounds on the potential then leads to the desired convergence time.

Lemma 4.21. *In any round of the algorithm in which the solution x^0 at the beginning of the round is feasible:*

$$\sum_{\left\{j:j\in S^- \wedge x_j^0 \geq \frac{\delta_j}{1-\beta}\right\}} x_j^0 \sum_{i=1}^m y_i(x^0) A_{ij} \geq \frac{1}{2} \sum_{j\in S^-} x_j^0 \sum_{i=1}^m y_i(x^0) A_{ij};$$

and

$$\sum_{\left\{j:j\in S^- \wedge x_j^0 \geq \frac{\delta_j}{1-\beta}\right\}} \left(x_j^0 \sum_{i=1}^m y_i(x^0) A_{ij} - (1+\gamma)w_j(x_j^0)^{1-\alpha}\right) \geq \frac{1}{2} \sum_{j\in S^-} \left(x_j^0 \sum_{i=1}^m y_i(x^0) A_{ij} - (1+\gamma)w_j(x_j^0)^{1-\alpha}\right).$$

Proof. If $x_j^0 \geq \frac{\delta_j}{1-\beta}$, $\forall j$, there is nothing to prove, so assume that there exists at least one j with $x_j^0 < \frac{\delta_j}{1-\beta}$. The proof proceeds as follows. First, we show that for each j for which x_j decreases by a factor less than $(1-\beta)$ there exists at least one x_p that appears in at least one constraint i in which x_j appears and decreases by a factor $(1-\beta)$. We then proceed to show that x_p is in fact such that

$$x_p^0 \sum_{l=1}^m y_l(x^0) A_{lp} = \Omega(n) x_j^0 \sum_{l=1}^m y_l(x^0) A_{lj}$$

26

and

$$x_p^0 \sum_{l=1}^m y_l(x^0) A_{lp} - (1+\gamma) w_p(x_p^0)^{1-\alpha} = \Omega(n) \left(x_j^0 \sum_{l=1}^m y_l(x^0) A_{lj} - (1+\gamma) w_j(x_j^0)^{1-\alpha} \right).$$

This will then imply that the terms $x_p^0 \sum_{l=1}^m y_l(x^0) A_{lp}$ and $x_p^0 \sum_{l=1}^m y_l(x^0) A_{lp} - (1+\gamma) w_p(x_p^0)^{1-\alpha}$ dominate *the sum* of all the terms corresponding to x_j's with $A_{ij} \neq 0$ and $x_j < \frac{\delta_j}{1-\beta}$, thus completing the proof.

From Lemma 4.7, for each $j \in S^-$ with $x_j < \frac{\delta_j}{1-\beta}$ there exists at least one constraint i such that:

- $\sum_{k=1}^n A_{ik} x_k^0 > 1 - \frac{\varepsilon}{2}$, and

- $y_i(x^0) \geq \frac{\sum_{l=1}^m y_l(x^0) A_{lj}}{m A_{\max}} \Rightarrow y_i(x^0) > (1-\beta)^\alpha \frac{1}{m A_{\max}} \frac{w_j}{\delta_j{}^\alpha} \frac{(x_j^0)^\alpha \sum_{l=1}^m y_l(x^0) A_{lj}}{w_j}$.

Therefore, there exists at least one x_p with $A_{ip} \neq 0$ such that $A_{ip} x_p^0 > \frac{1-\frac{\varepsilon}{2}}{n}$, which further gives $A_{ip}(x_p^0)^\alpha > \frac{(1-\frac{\varepsilon}{2})^\alpha}{n^\alpha} \cdot A_{ip}{}^{1-\alpha} \geq \frac{(1-\frac{\varepsilon}{2})^\alpha}{n^\alpha} \cdot A_{\max}{}^{1-\alpha}$, where the last inequality follows from $1 \leq A_{ip} \leq A_{\max}$ and $\alpha > 1$. Combining the inequality for $A_{ip}(x_p^0)^\alpha$ with the inequality for $y_i(x^0)$ above:

$$(x_p^0)^\alpha \sum_{l=1}^m y_l(x^0) A_{lp} \geq (x_p^0)^\alpha A_{ip} y_i(x^0)$$

$$\geq \frac{(1-\frac{\varepsilon}{2})^\alpha}{n^\alpha} \cdot A_{\max}{}^{1-\alpha} (1-\beta)^\alpha \frac{1}{m A_{\max}} \frac{w_j}{\delta_j{}^\alpha} \frac{(x_j^0)^\alpha \sum_{l=1}^m y_l(x^0) A_{lj}}{w_j}$$

$$= C \cdot \frac{(1-\frac{\varepsilon}{2})^\alpha}{n^\alpha m A_{\max}{}^\alpha} (1-\beta)^\alpha \frac{(x_j^0)^\alpha \sum_{l=1}^m y_l(x^0) A_{lj}}{w_j} \quad \text{(from } C = \frac{w_j}{\delta_j^\alpha})$$

$$\geq 2n w_{\max} (1-\beta)^\alpha \left(1 - \frac{\varepsilon}{2} \right)^\alpha \frac{(x_j^0)^\alpha \sum_{l=1}^m y_l(x^0) A_{lj}}{w_j} \quad \text{(from } C \geq 2 w_{\max} n^{\alpha+1} m A_{\max}{}^{2\alpha-1}).$$

Using the generalized Bernoulli's inequality: $\left(1 - \frac{\varepsilon}{2} \right)^\alpha > 1 - \frac{\varepsilon\alpha}{2}$ and $(1-\beta)^\alpha > (1 - \beta\alpha)$ [37], and recalling that $\varepsilon\alpha \leq \frac{9}{10}$, $\beta \leq \frac{\gamma\varepsilon}{5} = \frac{\varepsilon^2}{20} \leq \frac{\varepsilon}{120}$, we further get:

$$(x_p^0)^\alpha \sum_{l=1}^m y_l(x^0) A_{lp} \geq 2n w_{\max} \left(1 - \frac{9}{10 \cdot 120} \right) \left(1 - \frac{9}{20} \right) \cdot \frac{(x_j^0)^\alpha \sum_{l=1}^m y_l(x^0) A_{lj}}{w_j}$$

$$\geq n w_{\max} \frac{(x_j^0)^\alpha \sum_{l=1}^m y_l(x^0) A_{lj}}{w_j},$$

which further implies:

$$\frac{(x_p^0)^\alpha \sum_{l=1}^m y_l(x^0) A_{lp}}{w_p} \geq n \cdot \frac{(x_j^0)^\alpha \sum_{l=1}^m y_l(x^0) A_{lj}}{w_j}, \tag{50}$$

as $w_p \leq w_{\max}$. Since x_j decreases, $\frac{(x_j^0)^\alpha \sum_{l=1}^m y_l(x^0) A_{lj}}{w_j} \geq 1 + \gamma$, and therefore x_p decreases as well.

Using similar arguments, as $A_{ip} x_p^0 > \frac{1-\frac{\varepsilon}{2}}{n}$ and recalling that $y_i(x^0) \geq \frac{1}{m A_{\max}} \sum_{l=1}^m A_{lj} y_l(x^0) > \frac{1}{m A_{\max}} \frac{1-\beta}{\delta_j} \cdot x_j^0 \sum_{l=1}^m A_{lj} y_l(x^0)$:

$$x_p^0 \sum_{l=1}^m y_l(x^0) A_{lp} \geq x_p^0 A_{ip} y_i(x^0) \geq \frac{1-\frac{\varepsilon}{2}}{n} \frac{1}{m A_{\max}} \frac{1-\beta}{\delta_j} \cdot x_j^0 \sum_{l=1}^m A_{lj} y_l(x^0)$$

$$\geq n x_j^0 \sum_{l=1}^m A_{lj} y_l(x^0), \tag{51}$$

27

as $\delta_j \leq \frac{1}{2^{1/\alpha}n^2 mA_{\max}}$ and $2^{1/\alpha}(1 - \frac{\varepsilon}{2})(1 - \beta) \geq 2^{\frac{10}{9}\varepsilon}(1 - \frac{\varepsilon}{2})(1 - \frac{\varepsilon^2}{20}) \geq 1$ (since $\varepsilon \in (0, 1/6]$).

From (51), it follows that

$$x_p^0 \sum_{l=1}^m y_l(x^0)A_{lp} \geq \sum_{\{k\in S^-:x_k<\frac{\delta_k}{1-\beta}\wedge A_{ik}\neq 0\}} x_k^0 \sum_{l=1}^m y_l(x^0)A_{lk},$$

which further implies the first part of the lemma.

For the second part, consider the following two cases:

Case 1: $w_p(x_p^0)^{1-\alpha} \geq w_j(x_j^0)^{1-\alpha}$. Then:

$$x_p^0 \sum_{l=1}^m y_l(x^0)A_{lp} - (1+\gamma)w_p(x_p^0)^{1-\alpha} = w_p(x_p^0)^{1-\alpha}\left(\frac{(x_p^0)^\alpha \sum_{l=1}^m y_l(x^0)A_{lp}}{w_p} - (1+\gamma)\right)$$

$$\geq w_j(x_j^0)^{1-\alpha}\left(\frac{(x_p^0)^\alpha \sum_{l=1}^m y_l(x^0)A_{lp}}{w_p} - (1+\gamma)\right)$$

$$\geq w_j(x_j^0)^{1-\alpha}\left(n\frac{(x_j^0)^\alpha \sum_{l=1}^m y_l(x^0)A_{lj}}{w_j} - (1+\gamma)\right) \quad \text{(from (50))}$$

$$\geq nw_j(x_j^0)^{1-\alpha}\left(\frac{(x_j^0)^\alpha \sum_{l=1}^m y_l(x^0)A_{lj}}{w_j} - (1+\gamma)\right)$$

$$= n\left(x_j^0 \sum_{l=1}^m y_l(x^0)A_{lj} - (1+\gamma)w_j(x_j^0)^{1-\alpha}\right),$$

implying the second part of the lemma.

Case 2: $w_p(x_p^0)^{1-\alpha} < w_j(x_j^0)^{1-\alpha}$. Then:

$$x_p^0 \sum_{l=1}^m y_l(x^0)A_{lp} - (1+\gamma)w_p(x_p^0)^{1-\alpha} > x_p^0 \sum_{l=1}^m y_l(x^0)A_{lp} - (1+\gamma)w_j(x_j^0)^{1-\alpha}$$

$$\geq nx_j^0 \sum_{l=1}^m y_l(x^0)A_{lj} - (1+\gamma)w_j(x_j^0)^{1-\alpha} \quad \text{(from (51))}$$

$$\geq n\left(x_j^0 \sum_{l=1}^m y_l(x^0)A_{lj} - (1+\gamma)w_j(x_j^0)^{1-\alpha}\right),$$

thus implying the second part of the lemma and completing the proof. $\qquad\square$

The following lemma lower-bounds the increase in the potential, in each round.

Lemma 4.22. *Let x^0 and x^1 denote the values of x before and after any fixed round, respectively, and let $S^+ = \{j : x_j^1 > x_j^0\}$, $S^- = \{j : x_j^1 < x_j^0\}$. The potential increase in the round is lower bounded as:*

1. $\Phi(x^1) - \Phi(x^0) \geq \Omega(\beta\gamma) \sum_{j\in\{S^+\cup S^-\}} x_j^0 \sum_{i=1}^m y_i(x^0)A_{ij}$;

2. $\Phi(x^1) - \Phi(x^0) \geq \Omega\left(\frac{\beta}{(1-\beta)^\alpha}\right)\left(\sum_{j=1}^n x_j^0 \sum_{i=1}^m y_i(x^0) - (1+\gamma)\sum_{j=1}^n w_j(x_j^0)^{1-\alpha}\right)$;

3. $\Phi(x^1) - \Phi(x^0) \geq \Omega\left(\frac{\beta}{(1+\beta)^\alpha}\right)\left((1-\gamma)\sum_{j=1}^n w_j(x_j^0)^{1-\alpha} - \sum_{j=1}^n x_j^0 \sum_{i=1}^m y_i(x^0)\right)$.

Proof.

Proof of 1. From Lemma 4.9:

$$\Phi(x^1) - \Phi(x^0) \geq \sum_{j=1}^n w_j \frac{|x_j^1 - x_j^0|}{(x_j^1)^\alpha}\left|1 - \frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j}\right|.$$

Let $\xi_j(x^1) = \frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j}$. From the proof of Lemma 4.9, if $x_j^1 - x_j^0 > 0$, then $1 - \xi_j(x^1) \geq$

28

$\frac{3}{4}\gamma \geq \frac{3}{4}\gamma\xi_j(x^1)$, as $0 < \xi_j(x^1) \leq 1 - \frac{3}{4}\gamma$. If $x_j^1 - x_j^0 < 0$, then $1 - \xi_j(x^1) \leq -\frac{\gamma}{2}$, which implies $1 \leq \xi_j(x^1)(1 + \frac{\gamma}{2})^{-1}$, and thus $1 - \xi_j(x^1) \leq \xi_j(x^1)((1 + \gamma/2)^{-1} - 1) = \xi_j(x^1)\frac{-\gamma/2}{1+\gamma/2} < -\xi_j(x^1)\frac{\gamma/2}{3/2} = -\frac{\gamma}{3}\xi_j(x^1)$. Therefore: $|1 - \xi_j(x^1)| \geq \frac{\gamma}{3}\xi_j(x^1) \Leftrightarrow \left|1 - \frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j}\right| \geq \frac{\gamma}{3}\frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j}$, which further gives:

$$\Phi(x^1) - \Phi(x^0) \geq \sum_{j=1}^n w_j \frac{|x_j^1 - x_j^0|}{(x_j^1)^\alpha} \frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j} \geq \frac{\gamma}{3} \sum_{j=1}^n |x_j^1 - x_j^0| \cdot \sum_{i=1}^m y_i(x^1)A_{ij}.$$

If $j \in S^+$, then $x_j^1 = (1+\beta)x_j^0$, and therefore $|x_j^1 - x_j^0| \cdot \sum_{i=1}^m y_i(x^1)A_{ij} = \left(1 - \frac{1}{1+\beta}\right) x_j^1 \sum_{i=1}^m y_i(x^1)A_{ij} \geq \left(1 - \frac{\gamma}{4}\right) \frac{\beta}{1+\beta} x_j^0 \sum_{i=1}^m y_i(x^0)A_{ij}$.

Similarly, if $j \in S^-$ and $x_j^0 \geq \frac{\delta_j}{1-\beta}$, then $x_j^1 = (1 - \beta)x_j^0$, and therefore $|x_j^1 - x_j^0| \cdot \sum_{i=1}^m y_i(x^1)A_{ij} = \left(\frac{1}{1-\beta} - 1\right) x_j^1 \sum_{i=1}^m y_i(x^1)A_{ij} \geq \left(1 - \frac{\gamma}{4}\right) \frac{\beta}{1-\beta} x_j^0 \sum_{i=1}^m y_i(x^0)A_{ij}$. Using part 1 of Lemma 4.21:

$$\Phi(x^1) - \Phi(x^0) \geq \frac{\gamma}{6} \frac{\beta}{1+\beta} \sum_{j\in\{S^+\cup S^-\}} x_j^0 \sum_{i=1}^m y_i(x^0)A_{ij}.$$

Proof of 2: Consider $j \in S^-$ such that $x_j^0 \geq \frac{\delta_j}{1-\beta}$. Then $x_j^1 = (1 - \beta)x_j^0$, $\frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j} \geq (1 + \gamma)$, and using Lemma 4.9:

$$\Phi(x^1) - \Phi(x^0) \geq \sum_{\{j\in S^-: x_j^0 \geq \frac{\delta_j}{1-\beta}\}} w_j \frac{|x_j^1 - x_j^0|}{(x_j^1)^\alpha} \left|1 - \frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j}\right|$$

$$\geq \sum_{\{j\in S^-: x_j^0 \geq \frac{\delta_j}{1-\beta}\}} w_j \frac{\beta}{(1-\beta)^\alpha}(x_j^0)^{1-\alpha} \left(\frac{(x_j^1)^\alpha \sum_{i=1}^m y_i(x^1)A_{ij}}{w_j} - 1\right)$$

$$\geq \frac{\beta}{(1-\beta)^\alpha} \sum_{\{j\in S^-: x_j^0 \geq \frac{\delta_j}{1-\beta}\}} w_j(x_j^0)^{1-\alpha} \left((1 - \gamma/4)\frac{(x_j^0)^\alpha \sum_{i=1}^m y_i(x^0)A_{ij}}{w_j} - 1\right)$$

$$\geq (1 - \gamma/4)\frac{\beta}{(1-\beta)^\alpha} \sum_{\{j\in S^-: x_j^0 \geq \frac{\delta_j}{1-\beta}\}} w_j(x_j^0)^{1-\alpha} \left(\frac{(x_j^0)^\alpha \sum_{i=1}^m y_i(x^0)A_{ij}}{w_j} - (1 + \gamma)\right)$$

$$= (1 - \gamma/4)\frac{\beta}{(1-\beta)^\alpha} \sum_{\{j\in S^-: x_j^0 \geq \frac{\delta_j}{1-\beta}\}} \left(x_j^0 \sum_{i=1}^m y_i(x^0) - (1 + \gamma)w_j(x_j^0)^{1-\alpha}\right).$$

Using the second part of Lemma 4.21 and the fact that for $k \notin S^-$: $\frac{(x_k^0)^\alpha \sum_{i=1}^m y_i(x^0)A_{ik}}{w_k} < (1 + \gamma)$, we get the desired result:

$$\Phi(x^1) - \Phi(x^0) \geq \frac{1}{2}(1 - \gamma/4)\frac{\beta}{(1-\beta)^\alpha} \left(\sum_{j=1}^n x_j^0 \sum_{i=1}^m y_i(x^0) - (1 + \gamma)\sum_{j=1}^n w_j(x_j^0)^{1-\alpha}\right).$$

Proof of 3: The proof is equivalent to the proof of Lemma 4.10, part 2, and is omitted for brevity. $\qquad\square$

Consider the following definition of a stationary round:

Definition 4.23. *(Stationary round.) A round is stationary, if both:*

1. $\sum_{j\in\{S^+\cup S^-\}} x_j^0 \sum_{i=1}^m y_i(x)A_{ij} \leq \gamma \sum_{j=1}^n w_j(x_j^0)^{1-\alpha}$, *and*

29

2. $(1 - 2\gamma) \sum_{j=1}^{n} w_j (x_j^0)^{1-\alpha} \leq \sum_{j=1}^{n} x_j^0 \sum_{i=1}^{m} y_i(x^0) A_{ij}$

hold, where $S^+ = \{j : x_j^1 > x_j^0\}$, $S^- = \{j : x_j^1 < x_j^0\}$. *Otherwise, the round is non-stationary.*

The following two technical propositions are used in Lemma 4.26 for bounding the duality gap in stationary rounds.

Proposition 4.24. *After the initial the initial $\tau_0 + \tau_1$ rounds, where $\tau_0 = \frac{1}{\beta} \ln(1/\delta_{\min})$, $\tau_1 = \frac{1}{\beta} \ln(n A_{\max})$, it is always true that $G_\alpha(x, y(x)) \leq \sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{\alpha-1} \left(1 + (1 + 3\varepsilon)(\alpha - 1)\xi_j - \alpha \xi_j^{\frac{\alpha-1}{\alpha}}\right)$, where $\xi_j = \frac{x_j^\alpha \sum_i y_i(x) A_{ij}}{w_j}$.*

Proof. Recall from (2) that the duality gap for x, y in (P_α) is given as:

$$G_\alpha(x, y) = \sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{1-\alpha} \left(\left(\frac{w_j}{x_j^\alpha \sum_{i=1}^{m} y_i A_{ij}} \right)^{\frac{1-\alpha}{\alpha}} - 1 \right) + \sum_{i=1}^{m} y_i - \sum_{j=1}^{n} w_j x_j^{1-\alpha} \cdot \left(\frac{x_j^\alpha \sum_{j=1}^{n} A_{ij} y_i}{w_j} \right)^{\frac{\alpha-1}{\alpha}}.$$

From Lemma 4.6, after at most initial $\tau_0 + \tau_1$ rounds:

$$\sum_{i=1}^{m} y_i \leq (1 + 3\varepsilon) \sum_{j=1}^{n} x_j \sum_{i=1}^{m} y_i A_{ij}$$

$$= (1 + 3\varepsilon) \sum_{j=1}^{n} w_j x_j^{1-\alpha} \cdot \left(\frac{x_j^\alpha \sum_{i=1}^{m} y_i A_{ij}}{w_j} \right),$$

and letting $\xi_j = \frac{x_j^\alpha \sum_{i=1}^{m} y_i A_{ij}}{w_j}$, we get:

$$G_\alpha(x, y) \leq \sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{1-\alpha} \left(\xi_j^{\frac{\alpha-1}{\alpha}} - 1 + (1 + 3\varepsilon)(1 - \alpha)\xi_j - (1 - \alpha)\xi_j^{\frac{\alpha-1}{\alpha}} \right)$$

$$= \sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{1-\alpha} \left(\alpha \xi_j^{\frac{\alpha-1}{\alpha}} + (1 + 3\varepsilon)(1 - \alpha)\xi_j - 1 \right)$$

$$= \sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{\alpha-1} \left(1 + (1 + 3\varepsilon)(\alpha - 1)\xi_j - \alpha \xi_j^{\frac{\alpha-1}{\alpha}} \right).$$

$\square$

Proposition 4.25. *Let $r_\alpha(\xi_j) = \left(1 + (1 + 3\varepsilon)(\alpha - 1)\xi_j - \alpha \xi_j^{\frac{\alpha-1}{\alpha}}\right)$, where $\xi_j = \frac{x_j^\alpha \sum_{i=1}^{m} y_i A_{ij}}{w_j}$. If $\alpha > 1$ and $\xi_j \in (1 - \gamma, 1 + \gamma) \ \forall j \in \{1, ..., n\}$, then $r_\alpha(\xi_j) \leq \varepsilon(3\alpha - 2)$.*

Proof. Observe the first and the second derivative of $r_\alpha(\xi_j)$:

$$\frac{dr_\alpha(\xi_j)}{d\xi_j} = (\alpha - 1)(1 + 3\varepsilon - \xi_j^{-1/\alpha});$$

$$\frac{d^2 r_\alpha(\xi_j)}{d\xi_j^2} = \frac{1}{\alpha}(\alpha - 1)\xi_j^{-1/\alpha - 1}.$$

As $\xi_j > 0$, $r(\xi_j)$ is convex for $\alpha > 1$, and therefore: $r(\xi_j) \leq \max\{r(1 - \gamma), r(1 + \gamma)\}$. We have that:

$$r(1 - \gamma) = r(1 - \varepsilon/4) = 1 - \left(1 - \frac{\varepsilon}{4}\right)\left((1 - \alpha)(1 + 3\varepsilon) + \alpha(1 - \varepsilon/4)^{-1/\alpha}\right)$$

$$\leq 1 - \left(1 - \frac{\varepsilon}{4}\right)\left(1 - \alpha + 3\varepsilon(1 - \alpha) + \alpha(1 + \varepsilon/4)^{1/\alpha}\right)$$

$$\leq 1 - \left(1 - \frac{\varepsilon}{4}\right)\left(1 + \varepsilon/4 + 3\varepsilon(1 - \alpha)\right) \quad \text{(from } (1 + \varepsilon/4)^{1/\alpha} \geq 1 + \varepsilon/(4\alpha))$$

$$= 1 - 1 - \frac{\varepsilon}{4} + 3\varepsilon(\alpha - 1) + \frac{\varepsilon}{4}(1 + \varepsilon/4 - 3\varepsilon(\alpha - 1))$$

30

$$= \frac{\varepsilon^2}{16} + 3\varepsilon(\alpha - 1)\left(1 - \frac{\varepsilon}{4}\right)$$

$$\leq \varepsilon(3\alpha - 2).$$

On the other hand:

$$r(1+\gamma) = r(1+\varepsilon/4) = 1 - \left(1 + \frac{\varepsilon}{4}\right)\left((1-\alpha)(1+3\varepsilon) + \alpha(1+\varepsilon/4)^{-1/\alpha}\right)$$

$$\leq 1 - \left(1 + \frac{\varepsilon}{4}\right)\left(1 - \alpha + 3\varepsilon - 3\varepsilon\alpha + \alpha(1-\varepsilon/4)^{1/\alpha}\right)$$

$$\leq 1 - \left(1 + \frac{\varepsilon}{4}\right)\left(1 + \frac{11}{4}\varepsilon - 3\varepsilon\alpha\right)$$

$$\leq 1 - \left(1 + \frac{11}{4}\varepsilon - 3\varepsilon\alpha\right)$$

$$\leq \varepsilon(3\alpha - 2),$$

completing the proof. $\qquad\square$

The following lemma states that in any stationary round current solution is an $(1+\varepsilon(4\alpha-1))$-approximate solution.

Lemma 4.26. *In any stationary round that happens after the initial the initial $\tau_0 + \tau_1$ rounds, where $\tau_0 = \frac{1}{\beta}\ln(1/\delta_{\min})$, $\tau_1 = \frac{1}{\beta}\ln(nA_{\max})$, we have that $p_\alpha(x^*) - p_\alpha(x) \leq \varepsilon(4\alpha - 1)(-p_\alpha(x))$, where x^* is the optimal solution to (P_α) and x is the solution at the beginning of the round.*

Proof. Observe that for any $k \notin \{S^+ \cup S^-\}$ (by the definition of S^+ and S^-) we have that $1 - \gamma < \frac{x_k^\alpha \sum_{i=1}^m y_i(x)A_{ik}}{w_k} < 1 + \gamma$, which is equivalent to:

$$(1-\gamma)w_k x_k^{1-\alpha} < x_k \sum_{i=1}^m y_i(x)A_{ik} < (1+\gamma)w_k x_k^{1-\alpha} \quad \forall k \notin \{S^+ \cup S^-\}. \tag{52}$$

Using stationarity and (52):

$$(1-2\gamma)\sum_{j=1}^n w_j x_j^{1-\alpha} \leq \sum_{j=1}^n x_j \sum_{i=1}^n y_i(x)A_{ij}$$

$$= \sum_{l \in \{S^+\cup S^-\}} x_l \sum_{i=1}^n y_i(x)A_{il} + \sum_{k \notin \{S^+\cup S^-\}} x_k \sum_{i=1}^n y_i(x)A_{ik}$$

$$< \gamma \sum_{j=1}^n w_j x_j^{1-\alpha} + (1+\gamma)\sum_{k \notin \{S^+\cup S^-\}} w_k x_k^{1-\alpha}. \tag{53}$$

Since $\sum_{l \in \{S^+\cup S^-\}} w_l x_l^{1-\alpha} = \sum_{j=1}^n w_j x_j^{1-\alpha} - \sum_{k \notin \{S^+\cup S^-\}} w_k x_k^{1-\alpha}$, using (53):

$$(1-2\gamma)\sum_{l \in \{S^+\cup S^-\}} w_l x_l^{1-\alpha} < \gamma\sum_{j=1}^n w_j x_j^{1-\alpha} + (1+\gamma)\sum_{k \notin \{S^+\cup S^-\}} w_k x_k^{1-\alpha} - (1-2\gamma)\sum_{k \notin \{S^+\cup S^-\}} w_k x_k^{1-\alpha}$$

$$= \gamma\sum_{j=1}^n w_j x_j^{1-\alpha} + 3\gamma\sum_{k \notin \{S^+\cup S^-\}} w_k x_k^{1-\alpha}$$

$$\leq 4\gamma\sum_{j=1}^n w_j x_j^{1-\alpha},$$

and therefore:

$$\sum_{l \in \{S^+\cup S^-\}} w_l x_l^{1-\alpha} < \frac{4\gamma}{1-2\gamma}\sum_{j=1}^n w_j x_j^{1-\alpha} < 5\gamma\sum_{j=1}^n w_j x_j^{1-\alpha}, \tag{54}$$

31

as $\gamma = \frac{\varepsilon}{4}$ and $\varepsilon \leq \frac{1}{6}$.

As $p_\alpha(x^*) - p_\alpha(x) \leq G(x, y(x))$, from Proposition 4.24:

$$p_\alpha(x^*) - p_\alpha(x) \leq \sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{\alpha - 1} \left(1 + (1 + 3\varepsilon)(\alpha - 1)\xi_j - \alpha\xi_j^{\frac{\alpha-1}{\alpha}} \right)$$

$$= \sum_{k \notin \{S^+ \cup S^-\}} w_k \frac{x_k^{1-\alpha}}{\alpha - 1} \left(1 + (1 + 3\varepsilon)(\alpha - 1)\xi_k - \alpha\xi_k^{\frac{\alpha-1}{\alpha}} \right)$$

$$+ \sum_{l \in \{S^+ \cup S^-\}} w_l \frac{x_l^{1-\alpha}}{\alpha - 1} \left(1 + (1 + 3\varepsilon)(\alpha - 1)\xi_l - \alpha\xi_l^{\frac{\alpha-1}{\alpha}} \right).$$

From Proposition 4.25:

$$\sum_{k \notin \{S^+ \cup S^-\}} w_k \frac{x_k^{1-\alpha}}{\alpha - 1} \left(1 + (1 + 3\varepsilon)(\alpha - 1)\xi_k - \alpha\xi_k^{\frac{\alpha-1}{\alpha}} \right) \leq \varepsilon(3\alpha - 2) \sum_{k \notin \{S^+ \cup S^-\}} w_k \frac{x_k^{1-\alpha}}{\alpha - 1}$$

$$\leq \varepsilon(3\alpha - 2) \sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{\alpha - 1}$$

$$= \varepsilon(3\alpha - 2)(-p_\alpha(x)). \tag{55}$$

Observe $\sum_{l \in \{S^+ \cup S^-\}} w_l \frac{x_l^{1-\alpha}}{\alpha-1} \left(1 + (1 + 3\varepsilon)(\alpha - 1)\xi_l - \alpha\xi_l^{\frac{\alpha-1}{\alpha}} \right)$. Since $\alpha > 1$, each $w_l \frac{x_l^{1-\alpha}}{\alpha-1} > 0$, and therefore:

$$\sum_{l \in \{S^+ \cup S^-\}} w_l \frac{x_l^{1-\alpha}}{\alpha - 1} \left(1 + (1 + 3\varepsilon)(\alpha - 1)\xi_l - \alpha\xi_l^{\frac{\alpha-1}{\alpha}} \right)$$

$$\leq \sum_{l \in \{S^+ \cup S^-\}} w_l \frac{x_l^{1-\alpha}}{\alpha - 1} \left((1 + 3\varepsilon)(\alpha - 1)\xi_l + 1 \right)$$

$$= \sum_{l \in \{S^+ \cup S^-\}} w_l \frac{x_l^{1-\alpha}}{\alpha - 1} \left((1 + 3\varepsilon)(\alpha - 1)\frac{x_l^\alpha \sum_{i=1}^{m} y_i(x)A_{il}}{w_l} + 1 \right)$$

$$= (1 + 3\varepsilon) \sum_{l \in \{S^+ \cup S^-\}} x_l \sum_{i=1}^{m} y_i(x)A_{il} + \sum_{l \in \{S^+ \cup S^-\}} w_l \frac{x_l^{1-\alpha}}{\alpha - 1}.$$

Now, from stationarity $\sum_{l \in \{S^+ \cup S^-\}} x_l \sum_{i=1}^{m} y_i(x)A_{il} < \gamma \sum_{j=1}^{n} w_j x_j^{1-\alpha}$ and using (54) we get:

$$\sum_{l \in \{S^+ \cup S^-\}} w_l \frac{x_l^{1-\alpha}}{\alpha - 1} \left(1 + (1 + 3\varepsilon)(\alpha - 1)\xi_j - \alpha\xi_j^{\frac{\alpha-1}{\alpha}} \right) < \sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{\alpha - 1} (\gamma(1 + 3\varepsilon)(\alpha - 1) + 5\gamma)$$

$$\leq -p_\alpha(x) \left(\frac{3\varepsilon}{8}\alpha + \varepsilon \right). \tag{56}$$

Finally, combining (55) and (56): $p_\alpha(x^*) - p_\alpha(x) < \varepsilon(4\alpha - 1)(-p_\alpha(x))$. $\qquad \square$

The following two lemmas are used for lower-bounding the potential increase in non-stationary rounds.

Lemma 4.27. *Consider any non-stationary round that happens after the initial $\tau_0 + \tau_1$ rounds, where $\tau_0 = \frac{1}{\beta}\ln(1/\delta_{\min})$, $\tau_1 = \frac{1}{\beta}\ln(nA_{\max})$. Let x^0 and x^1 denote the values of x before and after the round update. If $\frac{1}{\kappa}\sum_i y_(x^0) \geq -\sum_j w_j \frac{(x_j^0)^{1-\alpha}}{1-\alpha}$, then $\Phi(x^1) - \Phi(x^0) \geq \Omega(\gamma^3)(-\Phi(x^0))$.*

Proof. Observe that as $\frac{1}{\kappa}\sum_i y_(x^0) \geq -\sum_j w_j \frac{(x_j^0)^{1-\alpha}}{1-\alpha}$,

$$-\Phi(x_0) \leq 2 \cdot \frac{1}{\kappa}\sum_i y_(x^0) \leq \frac{2(1-3\varepsilon)}{\kappa}\sum_{j=1}^n x_j^0 \sum_{i=1}^m y_i(x^0)A_{ij},$$

where the last inequality follows from Lemma 4.6.

Since the round is not stationary, we have that either:

1. $\sum_{j \in S^- \cup S^+} x_j^0 \sum_i y_i(x)A_{ij} > \gamma \sum_{j=1}^n w_j(x_j^0)^{1-\alpha}$, or

2. $(1-2\gamma)\sum_{j=1}^n w_j(x_j^0)^{1-\alpha} > \sum_{j=1}^n x_j^0 \sum_{i=1}^m y_i(x^0)A_{ij}$.

Case 1: $\sum_{j \in S^- \cup S^+} x_j^0 \sum_i y_i(x)A_{ij} > \gamma \sum_{j=1}^n w_j(x_j^0)^{1-\alpha}$. If:

$$\sum_{j=1}^n x_j^0 \sum_{i=1}^m y_i(x^0) \leq (1+2\gamma)\sum_{j=1}^n w_j(x_j^0)^{1-\alpha},$$

then

$$\sum_{j \in S^- \cup S^+} x_j^0 \sum_i y_i(x)A_{ij} > \frac{\gamma}{1+2\gamma}\sum_{j=1}^n x_j^0 \sum_{i=1}^m y_i(x^0)A_{ij} = \Omega(\gamma)\sum_{j=1}^n x_j^0 \sum_{i=1}^m y_i(x^0)A_{ij},$$

and, from the first part of Lemma 4.22, the potential increase is lower bounded as:

$$\begin{aligned}
\Phi(x^1) - \Phi(x^0) &\geq \Omega(\beta\gamma^2)\sum_{j=1}^n x_j^0 \sum_{i=1}^m y_i(x^0)A_{ij} \\
&= \Omega(\beta\kappa\gamma^2)(-\Phi(x^0)) \\
&= \Omega(\gamma^3)(-\Phi(x^0)).
\end{aligned}$$

On the other hand, if:

$$\sum_{j=1}^n x_j^0 \sum_{i=1}^m y_i(x^0) > (1+2\gamma)\sum_{j=1}^n w_j(x_j^0)^{1-\alpha},$$

then, from the second part of Lemma 4.22:

$$\begin{aligned}
\Phi(x^1) - \Phi(x^0) &\geq \Omega(\beta\gamma)\sum_{j=1}^n x_j^0 \sum_{i=1}^m y_i(x^0)A_{ij} \\
&= \Omega(\beta\gamma\kappa)(-\Phi(x^0)) \\
&= \Omega(\gamma^2)(-\Phi(x^0)).
\end{aligned}$$

Case 2: $(1-2\gamma)\sum_{j=1}^n w_j(x_j^0)^{1-\alpha} > \sum_{j=1}^n x_j^0 \sum_{i=1}^m y_i(x^0)A_{ij}$. Then, using the third part of Lemma 4.22:

$$\begin{aligned}
\Phi(x^1) - \Phi(x^0) &\geq \Omega\left(\frac{\beta}{(1+\beta)^\alpha}\gamma\right)\sum_{j=1}^n x_j^0 \sum_{i=1}^m y_i(x^0)A_{ij} \\
&= \Omega\left(\beta\gamma\right)\sum_{j=1}^n x_j^0 \sum_{i=1}^m y_i(x^0)A_{ij} \\
&= \Omega(\beta\gamma\kappa)(-\Phi(x^0)) \\
&= \Omega(\gamma^2)(-\Phi(x^0)),
\end{aligned}$$

where in the second line we have used that $\frac{\beta}{(1+\beta)^\alpha} = \Theta(\beta)$. This can be shown using the generalized Bernoulli's inequality and $\varepsilon\alpha \leq \frac{9}{10}$ as follows:

$$\frac{1}{(1+\beta)^\alpha} \geq (1-2\beta)^\alpha \geq 1 - 2\alpha\beta = 1 - \frac{\alpha}{k+\alpha}\cdot\frac{\varepsilon}{10} \geq 1 - \frac{9}{100} = \Theta(1).$$

□

Lemma 4.28. *Consider any non-stationary round that happens after the initial $\tau_0 + \tau_1$ rounds, where $\tau_0 = \frac{1}{\beta} \ln(1/\delta_{\min})$, $\tau_1 = \frac{1}{\beta} \ln(nA_{\max})$. Let x^0 and x^1 denote the values of x before and after the round update. If $\frac{1}{\kappa} \sum_i y_(x^0) < - \sum_j w_j \frac{(x_j^0)^{1-\alpha}}{1-\alpha}$, then $\Phi(x^1) - \Phi(x^0) \geq \Omega\left(\beta\gamma^2\right)(\alpha - 1)(-\Phi(x^0))$.*

Proof. Observe that as $\frac{1}{\kappa} \sum_i y_(x^0) < - \sum_j w_j \frac{(x_j^0)^{1-\alpha}}{1-\alpha}$,

$$-\Phi(x_0) \leq -2 \sum_j w_j \frac{(x_j^0)^{1-\alpha}}{1 - \alpha} = \frac{2}{\alpha - 1} \sum_j w_j (x_j^0)^{1-\alpha}.$$

From the definition of a stationary round, we have either of the following two cases:

Case 1: $\sum_{j \in \{S^+ \cup S^-\}} x_j \sum_{i=1}^m y_i(x) A_{ij} > \gamma \sum_{j=1}^n w_j x_j^{1-\alpha}$. From the first part of Lemma 4.22, the increase in the potential is: $\Phi(x^1) - \Phi(x^0) \geq \Omega\left(\beta\gamma^2\right) \sum_{j=1}^n w_j x_j^{1-\alpha}$. As $-\Phi(x^0) \leq \frac{2}{\alpha-1} \sum_j w_j (x_j^0)^{1-\alpha}$, the increase in the potential is at least:

$$\Phi(x^1) - \Phi(x^0) \geq \Omega(\beta\gamma^2)(\alpha - 1)(-\Phi(x^0)).$$

Case 2: $(1 - 2\gamma) \sum_{j=1}^n w_j x_j^{1-\alpha} > \sum_{j=1}^n x_j \sum_{i=1}^m y_i(x) A_{ij}$. Using part 3 of Lemma 4.22, the increase in the potential is then $\Phi(x^1) - \Phi(x^0) \geq \Omega\left(\frac{\beta}{(1+\beta)^\alpha}\gamma\right) \sum_{j=1}^n w_j x_j^{1-\alpha}$. Therefore, using that $\frac{\beta}{(1+\beta)^\alpha} = \Theta(\beta)$ as in the proof of Lemma 4.27:

$$\Phi(x^1) - \Phi(x^0) \geq \Omega(\beta\gamma)(\alpha - 1)(-\Phi(x^0)).$$

□

Proof of Theorem 4.3. We will bound the total number of non-stationary rounds that happen after the initial $\tau_0 + \tau_1$ rounds, where $\tau_0 = \frac{1}{\beta} \ln(1/\delta_{\min})$, $\tau_1 = \frac{1}{\beta} \ln(nA_{\max})$. The total convergence time is then at most the sum of $\tau_0 + \tau_1$ rounds and the number of non-stationary rounds that happen after the initial $\tau_0 + \tau_1$ rounds, since, from Lemma 4.26, in any stationary round: $p(x^*) - p(x) \leq \varepsilon(4\alpha - 1)(-p(x))$.

Consider the non-stationary rounds that happen after the initial $\tau_0 + \tau_1$ rounds. As $x_j \in [\delta_j, 1], \forall j$, it is simple to show that:

$$\frac{W}{\alpha - 1} \leq \sum_j w_j \frac{x_j^{1-\alpha}}{\alpha - 1} \leq \frac{W}{\alpha - 1} \cdot 2R_w^{\frac{\alpha-1}{\alpha}} n^{2(\alpha-1)} m^{\alpha-1} A_{\max}^{2\alpha-1}, \tag{57}$$

and

$$0 < \frac{1}{\kappa} \sum_i y_i(x) \leq \frac{mC}{\kappa} \leq \varepsilon mC. \tag{58}$$

Recall that $\Phi(x) = - \sum_j w_j \frac{x_j^{1-\alpha}}{\alpha-1} - \frac{1}{\kappa} \sum_i y_i(x)$ and that the potential $\Phi(x)$ never decreases.

There can be two cases of non-stationary rounds: those in which $\sum_j w_j \frac{x_j^{1-\alpha}}{\alpha-1}$ dominates in the absolute value of the potential, and those in which $\frac{1}{\kappa} \sum_i y_i(x)$ dominates in the absolute value of the potential. We bound the total number of the non-stationary rounds in such cases as follows.

Case 1: $\frac{1}{\kappa} \sum_i y_i(x) \geq \sum_j w_j \frac{x_j^{1-\alpha}}{\alpha-1}$. From (57) and (58), in any such round, the negative potential is bounded as:

$$\Omega\left(\frac{W}{\alpha - 1}\right) \leq -\Phi(x) \leq O\left(\varepsilon mC\right).$$

Moreover, from Lemma 4.27, in each Case 1 non-stationary round, the potential increases by at least

34

$\Omega(\gamma^3)(-\Phi(x))$. It immediately follows that there can be at most:

$$O\left(\frac{1}{\gamma^3}\ln\left(\frac{\varepsilon mC}{\frac{W}{\alpha-1}}\right)\right) = O\left(\frac{1}{\gamma^3}\ln\left((\alpha-1)\varepsilon R_w nmA_{\max}\right)\right)$$

$$= O\left(\frac{1}{\varepsilon^3}\ln\left(R_w nmA_{\max}\right)\right) \qquad (59)$$

Case 1 non-stationary rounds, as $(\alpha-1)\varepsilon < \alpha\varepsilon \le \frac{9}{10}$.

Case 2: $\frac{1}{\kappa}\sum_i y_i(x) < \sum_j w_j \frac{x_j^{1-\alpha}}{\alpha-1}$. From (57) and (58), in any such round, the negative potential is bounded as:

$$\Omega\left(\frac{W}{\alpha-1}\right) \le -\Phi(x) \le O\left(\frac{W}{\alpha-1}\cdot R_w^{\frac{\alpha-1}{\alpha}}n^{2(\alpha-1)}m^{\alpha-1}A_{\max}^{2\alpha-1}\right).$$

Moreover, from Lemma 4.21, in each such non-stationary round the potential increases by at least $\Omega\left(\beta\gamma^2\right)(\alpha-1)(-\Phi(x^0))$. Therefore, there can be at most:

$$O\left(\frac{1}{\beta\gamma^2(\alpha-1)}\ln\left(\frac{\frac{W}{\alpha-1}\cdot R_w^{\frac{\alpha-1}{\alpha}}n^{2(\alpha-1)}m^{\alpha-1}A_{\max}^{2\alpha-1}}{\frac{W}{\alpha-1}}\right)\right) = O\left(\frac{1}{\beta\gamma^2}\ln(R_w^{\frac{1}{\alpha}}nmA_{\max})\right)$$

$$= O\left(\frac{1}{\varepsilon^4}\ln(R_w nmA_{\max})\ln\left(R_w\cdot\frac{nmA_{\max}}{\varepsilon}\right)\right) \qquad (60)$$

Case 2 non-stationary rounds.

The total number of initial $\tau_0 + \tau_1$ rounds can be bounded as:

$$\tau_0 + \tau_1 = \frac{1}{\beta}\ln(1/\delta_{\min}) + \frac{1}{\beta}\ln(nA_{\max})$$

$$= O\left(\frac{1}{\varepsilon^2}\ln\left(R_w nmA_{\max}\right)\ln\left(R_w\cdot\frac{nmA_{\max}}{\varepsilon}\right)\right). \qquad (61)$$

Combining (59), (60), and (61), the total convergence time is at most:

$$O\left(\frac{1}{\varepsilon^4}\ln\left(R_w\cdot nmA_{\max}\right)\ln\left(R_w\cdot\frac{nmA_{\max}}{\varepsilon}\right)\right).$$

Finally, running α-FAIRPSOLVER for the approximation parameter $\varepsilon' = \varepsilon/(4\alpha-1)$, we get that in any stationary round $p_\alpha(x^*) - p_\alpha(x) \le -\varepsilon p_\alpha(x)$, while the total number of non-stationary rounds is at most:

$$O\left(\frac{\alpha^4}{\varepsilon^4}\ln\left(R_w\cdot nmA_{\max}\right)\ln\left(R_w\cdot\frac{nmA_{\max}}{\varepsilon}\right)\right).$$

$\square$

4.4 Structural Properties of $\alpha-$Fair Allocations

Lower Bound on the Minimum Allocated Value. Recall (from Section 2) that the optimal solution x^* to (P_α) must lie in the positive orthant. We show in Lemma 4.29 that not only does x^* lie in the positive orthant, but the minimum element of x^* can be bounded below as a function of the problem parameters. This lemma motivates the choice of parameters δ_j in α-FAIRPSOLVER (Section 3).

Lemma 4.29. *Let $x^* = (x_1^*, ..., x_n^*)$ be the optimal solution to (P_α). Then $\forall j \in \{1, ..., n\}$:*

- $x_j^* \ge \left(\frac{w_j}{w_{\max}M}\min_{i:A_{ij}\ne 0}\frac{1}{n_i A_{ij}}\right)^{1/\alpha}$, *if $0 < \alpha \le 1$,*

- $x_j^* \ge A_{\max}^{(1-\alpha)/\alpha}\left(\frac{w_j}{w_{\max}M}\right)^{1/\alpha}\min_{i:A_{ij}\ne 0}\frac{1}{n_i A_{ij}}$, *if $\alpha > 1$,*

where $n_i = \sum_{j=1}^n \mathbb{1}_{\{A_{ij}\ne 0\}}$[7] is the number of non-zero elements in the ith row of the constraint matrix A,

---

[7]With the abuse of notation, $\mathbb{1}_{\{e\}}$ is the indicator function of the expression e, i.e., 1 if e holds, and 0 otherwise.

and $M = \min\{m, n\}$.

Proof. Fix α. Let:

$$\mu_j(\alpha) = \begin{cases} \left(\frac{w_j}{w_{\max} M} \min_{i:A_{ij} \neq 0} \frac{1}{n_i A_{ij}} \right)^{1/\alpha}, & \text{if } \alpha \leq 1 \\ A_{\max}^{(1-\alpha)/\alpha} \left(\frac{w_j}{w_{\max} M} \right)^{1/\alpha} \min_{i:A_{ij} \neq 0} \frac{1}{n_i A_{ij}}, & \text{if } \alpha > 1 \end{cases}.$$

For the purpose of contradiction, suppose that $x^* = (x_1^*, ..., x_n^*)$ is the optimal solution to (P_α), and $x_j^* < \mu_j(\alpha)$ for some fixed $j \in \{1, ..., n\}$.

To establish the desired result, we will need to introduce additional notation. We first break the set of (the indices of) constraints of the form $Ax \leq 1$ in which variable x_j appears with a non-zero coefficient into two sets, U and T:

- Let U denote the set of the constraints from (P_α) that are not tight at the given optimal solution x^*, and are such that $A_{u,j} \neq 0$ for $u \in U$. Let $s_u = 1 - \sum_{k=1}^n A_{uk} x_k$ denote the slack of the constraint $u \in U$.

- Let T denote the set of tight constraints from (P_α) that are such that $A_{tj} \neq 0$ for $t \in T$. Observe that since x^* is assumed to be optimal, $T \neq \emptyset$.

Let $\varepsilon_j = \min \left\{ \mu_j(\alpha) - x_j^*, \min_{u \in U} s_u / A_{uj} \right\}$. Notice that by increasing x_j to $x_j^* + \varepsilon_j$ none of the constraints from U can be violated (although all the constraints in T will; we deal with these violations in what follows).

In each constraint $t \in T$, there must exist at least one variable x_k such that $x_k^* > \dfrac{1}{n_t A_{tk}}$, because $\sum_{l=1}^n A_{tl} x_l^* = 1$, as each $t \in T$ is tight, and $x_j^* < \mu_j(\alpha) \leq \min_{i:A_{ij} \neq 0} \frac{1}{n_i A_{ij}} \leq \frac{1}{n_t A_{tj}}$. Select one such x_k in each constraint $t \in T$, and denote by K the set of indices of selected variables. Observe that $|K| \leq |T|$ ($\leq M$), since an x_k can appear in more than one constraint.

For each $k \in K$, let T_k denote the constraints in which x_k is selected, and let

$$\varepsilon_k = \max_{t \in T_k: A_{tk} \neq 0} \frac{A_{tj} \varepsilon_j}{A_{tk}}. \tag{62}$$

If we increase x_j by ε_j and decrease x_k by ε_k $\forall k \in K$, each of the constraints $t \in T$ will be satisfied since, from (62) and from the fact that only one x_k gets selected per constraint $t \in T$, $\varepsilon_j A_{tj} - \sum_{k \in K} \varepsilon_k A_{tk} \leq 0$. Therefore, to construct an alternative feasible solution x', we set $x_j' = x_j^* + \varepsilon_j$, $x_k' = x_k^* - \varepsilon_k$ for $k \in K$, and $x_l' = x_l^*$ for all the remaining coordinates $l \in \{1, ..., n\} \setminus (K \cup \{j\})$.

Since j is the only coordinate over which x gets increased in x', all the constraints $Ax' \leq 1$ are satisfied. For x' to be feasible, we must have in addition that $x_k' \geq 0$ for $k \in K$. We show that $x_k' = x_k^* - \varepsilon_k \geq 0$ as follows:

$$\varepsilon_k = \varepsilon_j \cdot \max_{t \in T_k: A_{tk} \neq 0} \frac{A_{tj}}{A_{tk}}$$

$$\leq \mu_j(\alpha) \cdot \max_{t \in T_k: A_{tk} \neq 0} \frac{A_{tj}}{A_{tk}}$$

$$\leq \min_{i:A_{ij} \neq 0} \frac{1}{n_i A_{ij}} \cdot \max_{t \in T_k: A_{tk} \neq 0} \frac{A_{tj}}{A_{tk}}$$

$$\leq \max_{t \in T_k: A_{tk} \neq 0} \frac{1}{n_t A_{tj}} \frac{A_{tj}}{A_{tk}}$$

$$\leq \max_{t \in T_k: A_{tk} \neq 0} \frac{1}{n_t A_{tk}}$$

$$< x_k^*,$$

36

where the second line follows from $\varepsilon_j \leq \mu_j(\alpha) - x_j^* \leq \mu_j(\alpha)$, and the last line follows from the choice of x_k.

The last part of the proof is to show that $\sum_{l=1}^{n} w_l \frac{x_l' - x_l^*}{x_l^{*\alpha}} > 0$, which contradicts the initial assumption that x^* is optimal, by the definition of α-fairness from Section 2. We have that:

$$
\begin{aligned}
\sum_{l=1}^{n} w_l \frac{x_l' - x_l^*}{x_l^{*\alpha}} &= w_j \frac{\varepsilon_j}{x_j^{*\alpha}} - \sum_{k \in K} w_k \frac{\varepsilon_k}{x_k^{*\alpha}} \\
&= \sum_{k \in K} \left(w_j \frac{\varepsilon_j}{x_j^{*\alpha}|K|} - w_k \frac{\varepsilon_k}{x_k^{*\alpha}} \right) \\
&= \sum_{k \in K} \left(\frac{w_j \varepsilon_j x_k^{*\alpha} - w_k \varepsilon_k x_j^{*\alpha}|K|}{x_j^{*\alpha} x_k^{*\alpha}|K|} \right).
\end{aligned}
\tag{63}
$$

Consider one term from the summation (63). From the choice of ε_k's, we know that for each ε_k there exist $t \in T$ such that $\varepsilon_k = \frac{\varepsilon_j A_{tj}}{A_{tk}}$, and at the same time (by the choice of x_k) we have $x_k^* > \frac{1}{n_t A_{tk}}$, so that

$$
w_j \varepsilon_j x_k^{*\alpha} > w_j \frac{\varepsilon_k A_{tk}}{A_{tj}} \left(\frac{1}{A_{tk} n_t} \right)^\alpha > \frac{w_k w_j \varepsilon_k}{w_{\max}} \frac{A_{tk}}{A_{tj}} \left(\frac{1}{A_{tk} n_t} \right)^\alpha.
\tag{64}
$$

Case 1. Suppose first that $\alpha \leq 1$. Then $x_k^{*\alpha} > \left(\frac{1}{A_{tk} n_t} \right)^\alpha \geq \frac{1}{A_{tk} n_t}$, as $A_{tk} \neq 0 \Rightarrow A_{tk} \geq 1$. Plugging into (64), we have:

$$
w_j \varepsilon_j x_k^{*\alpha} > \frac{w_k w_j \varepsilon_k}{w_{\max}} \frac{1}{n_t A_{tj}}.
\tag{65}
$$

By the initial assumption, $x_j^* < \mu_j(\alpha) = \left(\frac{w_j}{w_{\max} M} \min_{i:A_{ij} \neq 0} \frac{1}{n_i A_{ij}} \right)^{1/\alpha}$, and therefore

$$
w_k \varepsilon_k x_j^{*\alpha}|K| < \frac{w_k w_j \varepsilon_k}{w_{\max}} \frac{|K|}{M} \min_{i:A_{ij} \neq 0} \frac{1}{n_i A_{ij}} \leq \frac{w_k w_j \varepsilon_k}{w_{\max}} \frac{1}{n_t A_{tj}},
\tag{66}
$$

since it must be $|K| \leq M \, (= \min\{m, n\})$. From (65) and (66), we get that every term in the summation (63) is strictly positive, which implies:

$$
\sum_{l=1}^{n} w_l \frac{x_l' - x_l^*}{x_l^{*\alpha}} > 0,
$$

and therefore x^* is not optimal.

Case 2. Now suppose that $\alpha > 1$. Then

$$
x_j^* < \mu_j(\alpha) = A_{\max}^{(1-\alpha)/\alpha} \left(\frac{w_j}{w_{\max} M} \right)^{1/\alpha} \min_{i:A_{ij} \neq 0} \frac{1}{n_i A_{ij}} \leq A_{\max}^{(1-\alpha)/\alpha} \left(\frac{w_j}{w_{\max} M} \right)^{1/\alpha} \frac{1}{n_t A_{tj}}.
$$

Therefore:

$$
\begin{aligned}
w_k \varepsilon_k x_j^{*\alpha}|K| &< w_k \varepsilon_k \frac{w_j}{w_{\max} M} A_{\max}^{1-\alpha} \left(\frac{1}{n_t A_{tj}} \right)^\alpha |K| \\
&\leq w_k \frac{w_j}{w_{\max}} A_{\max}^{1-\alpha} \varepsilon_k \left(\frac{1}{A_{tk} n_t} \right)^\alpha \frac{A_{tk}^\alpha}{A_{tj}^\alpha} \\
&= w_k \frac{w_j}{w_{\max}} \frac{\varepsilon_k A_{tk}}{A_{tj}} \cdot \frac{(A_{tk}/A_{tj})^{\alpha-1}}{A_{\max}^{\alpha-1}} \left(\frac{1}{A_{tk} n_t} \right)^\alpha \\
&\leq w_k \frac{w_j}{w_{\max}} \frac{\varepsilon_k A_{tk}}{A_{tj}} \left(\frac{1}{A_{tk} n_t} \right)^\alpha,
\end{aligned}
\tag{67}
$$

as $|K| \leq M$, and $\frac{A_{tk}}{A_{tj}} \leq A_{\max}$ (since for any i, j: $1 \leq A_{ij} \leq A_{\max}$).

Finally, from (64) and (67) we get that every term in the summation (63) is positive, which yields a contradiction. □

Asymptotics of $\alpha-$Fair Allocations The following lemma states that for sufficiently small (but not too small) α, the values of the linear and the $\alpha-$fair objectives at their respective optimal solutions are approximately the same. This statement will then lead to a conclusion that to $\varepsilon-$approximately solve an $\alpha-$fair packing problem for a very small α, one can always use an $\varepsilon-$approximation packing LP algorithm.

Lemma 4.30. *Let (P_α) be an $\alpha-$fair packing problem with optimal solution x^*, and (P_0) be the LP with the same constraints and the same weights w as (P_α) and an optimal solution z^*. Then if $\alpha \leq \frac{\varepsilon/4}{\ln(nA_{\max}/\varepsilon)}$, we have that $\sum_j w_j z_j^* \geq (1 - 3\varepsilon) \sum_j \frac{(x_j^*)^{1-\alpha}}{1-\alpha}$, where $\varepsilon \in (0, 1/6]$.*

Proof. The proof outline is as follows. First, we show that the $\alpha-$fair objective $p_\alpha(x^*)$ can be upper-bounded by a linear objective as $p_\alpha(x^*) \equiv \sum_j w_j \frac{x_j^{*1-\alpha}}{1-\alpha} \leq (1 + O(\varepsilon)) \sum_j w_j x_j^*$. Then, to complete the proof, we use the optimality of z^* for the LP: $\sum_j w_j z_j^* \geq \sum_j w_j x_j^* (\geq (1 - O(\varepsilon)) \sum_j w_j \frac{x_j^{*1-\alpha}}{1-\alpha}$ from the first part of the proof).

Let $g(x_j) = \frac{x_j^{1-\alpha}}{1-\alpha} - (1 + \varepsilon)x_j$. Consider the case when $g(x_j) \leq 0$. Solving $g(x_j) \leq 0$ for x_j, we get that it should be

$$x_j \geq \left(\frac{1}{1-\alpha}\right)^{1/\alpha} \cdot \left(\frac{1}{1+\varepsilon}\right)^{1/\alpha}. \tag{68}$$

Choose α so that $\frac{1}{(1+\varepsilon)^{1/\alpha}} \leq \left(\frac{\varepsilon/4}{nA_{\max}}\right)$, which is equivalent to $\alpha \leq \frac{\ln(1+\varepsilon)}{\ln(4nA_{\max}/\varepsilon)}$. Then to have $g(x_j) \leq 0$, it suffices to have $x_j \geq \frac{\varepsilon}{nA_{\max}}$, because (i) $\left(\frac{1}{1-\alpha}\right)^{1/\alpha} \in [e, 4]$ for $\alpha \in [0, 1/2]$, where e is the base of the natural logarithm, and (ii) $\frac{1}{(1+\varepsilon)^{1/\alpha}} \leq \left(\frac{\varepsilon/4}{nA_{\max}}\right)$ by the choice of α.

Now, as $\alpha \leq \frac{\ln(1+\varepsilon)}{\ln(4nA_{\max}/\varepsilon)}$, summing over j such that $x_j^* \geq \frac{\varepsilon}{nA_{\max}}$ we have:

$$\sum_{j: x_j^* \geq \frac{\varepsilon}{nA_{\max}}} w_j \frac{(x_j^*)^{1-\alpha}}{1-\alpha} - (1 + \varepsilon) \sum_{j: x_j^* \geq \frac{\varepsilon}{nA_{\max}}} w_j x_j^* = \sum_{j: x_j^* \geq \frac{\varepsilon}{nA_{\max}}} w_j g(x_j^*) \leq 0 \tag{69}$$

Now we bound the rest of the terms in $p_\alpha(x^*)$, i.e., we consider $j : x_j^* < \frac{\varepsilon}{nA_{\max}}$. Observe that since $x_j = \frac{1}{nA_{\max}}$ for $j = \{1, ..., n\}$ is a feasible solution to (P_α) and x^* is the optimal solution to (P_α), we have that $\sum_j w_j \frac{(1/nA_{\max})^{1-\alpha}}{1-\alpha} \leq \sum_j w_j \frac{(x_j^*)^{1-\alpha}}{1-\alpha}$, which gives:

$$\sum_{j: x_j^* < \frac{\varepsilon}{nA_{\max}}} w_j \frac{(x_j^*)^{1-\alpha}}{1-\alpha} < \varepsilon^{1-\alpha} \sum_{j: x_j^* < \frac{\varepsilon}{nA_{\max}}} w_j \frac{(1/nA_{\max})^{1-\alpha}}{1-\alpha}$$

$$< \varepsilon^{1-\alpha} \sum_{j=1}^n w_j \frac{(x_j^*)^{1-\alpha}}{1-\alpha}$$

$$\leq 2\varepsilon \sum_{j=1}^n w_j \frac{(x_j^*)^{1-\alpha}}{1-\alpha}.$$

Therefore:

$$\sum_{j: x_j^* \geq \frac{\varepsilon}{nA_{\max}}} w_j \frac{(x_j^*)^{1-\alpha}}{1-\alpha} > (1 - 2\varepsilon) \sum_{j=1}^n w_j \frac{(x_j^*)^{1-\alpha}}{1-\alpha}. \tag{70}$$

38

Combining (69) and (70), we now get:

$$\sum_{j=1}^{n} w_j \frac{(x_j^*)^{1-\alpha}}{1-\alpha} < \frac{1+\varepsilon}{1-2\varepsilon} \cdot \sum_{j:x_j^* \geq \frac{\varepsilon}{nA_{\max}}} w_j x_j^*. \tag{71}$$

Finally, since z^* optimally solves (P_0) (which has the same constraints and weights as (P_α)), we have that x^* is feasible for (P_0), and using (71) and optimality of z^*, it follows that:

$$\sum_{j=1}^{n} w_j z_j^* \geq \sum_{j=1}^{n} w_j x_j^*$$

$$\geq \frac{1-2\varepsilon}{1+\varepsilon} \sum_{j=1}^{n} w_j \frac{(x_j^*)^{1-\alpha}}{1-\alpha}$$

$$\geq (1-3\varepsilon) \sum_{j=1}^{n} w_j \frac{(x_j^*)^{1-\alpha}}{1-\alpha},$$

as claimed. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Observing that for any $\alpha \in (0,1)$, $\frac{(z_j^*)^{1-\alpha}}{1-\alpha} \geq z_j^*$ (since, due to the scaling, $z_j^* \in [0,1]$), a simple corollary of Lemma 4.30 is that an $\varepsilon-$approximation z to (P_0) ($\sum_j w_j z_j \geq (1-\varepsilon) \sum_j w_j z_j^*$) is also an $O(\varepsilon)-$approximation to (P_α), for $\alpha \leq \frac{\varepsilon/4}{\ln(nA_{\max}/\varepsilon)}$. Thus, to find an $\varepsilon-$approximate solution for $\alpha \leq \frac{\varepsilon/4}{\ln(nA_{\max}/\varepsilon)}$, the packing LP algorithm of [7] can be run, which means that there is a stateless distributed algorithm that converges in $\mathrm{poly}(\ln(\varepsilon^{-1} R_w mn A_{\max})/\varepsilon)$ time for α arbitrarily close to zero.

The following two lemmas show that when α is sufficiently close to 1, (P_α) can be $\varepsilon-$approximated by $\varepsilon-$approximately solving (P_1) with the same constraints and weights.

Lemma 4.31. *Let x be an $\varepsilon-$approximate solution to a 1-fair packing problem (P_1) returned by α-*FAIRPSOLVER. *Then, for any $\alpha \in [1 - 1/\tau_0, 1)$, where $\tau_0 = \frac{1}{\beta} \ln(\frac{1}{\delta_{\min}})$, x is also a $2\varepsilon-$approximate solution to (P_α), where the only difference between (P_1) and (P_α) is in the value of α in the objective.*

Proof. Suppose that x is a solution in some stationary round, provided by α-FAIRPSOLVER run for $\alpha = 1$. Fix that round. It is clear that if x is feasible in (P_1), it is also feasible in (P_α), since all the constraints in (P_1) and (P_α) are the same by the initial assumption. All that is required for a dual solution y to be feasible is that $y_i \geq 0$, for all i, and therefore $y(x)$ is a feasible dual solution for (P_α). The rest of the proof follows by bounding the duality gap $G_\alpha(x, y(x))$. Recall from (2) that:

$$G_\alpha(x, y(x)) = \sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{1-\alpha} \left(\left(\frac{x_j^\alpha \sum_{i=1}^{m} y_i A_{ij}}{w_j} \right)^{\frac{\alpha-1}{\alpha}} - 1 \right) + \sum_{i=1}^{m} y_i - \sum_{j=1}^{n} w_j x_j^{1-\alpha} \cdot \left(\frac{x_j^\alpha \sum_{i=1}^{m} A_{ij} y_i}{w_j} \right)^{\frac{\alpha-1}{\alpha}}. \tag{72}$$

Since x is a solution from a stationary round, from the second part of the definition of a stationary round (Definition 4.17), we have that:

$$\sum_{j=1}^{n} x_j \sum_{i=1}^{n} y_i(x) A_{ij} \leq (1+2\gamma) \sum_{k=1}^{n} w_k.$$

Further, from Lemma 4.6:

$$\sum_{i=1}^{m} y_i(x) \leq (1+3\varepsilon) \sum_{j=1}^{n} x_j \sum_{i=1}^{n} y_i(x) A_{ij} \leq (1+3\varepsilon)(1+2\gamma) \sum_{k=1}^{n} w_k. \tag{73}$$

Next, we show that:

$$x_j^{1-\alpha} \geq 1 - \gamma, \quad \forall j. \tag{74}$$

39

Rearranging the terms and taking logarithms of both sides in (74), we obtain the equivalent inequality $1 - \alpha \leq \frac{\ln(1/(1-\gamma))}{\ln(1/x_j)}$. Recall from α-FAIRPSOLVER that in every (except for, maybe, the first) round $x_j \geq \delta_j \geq \delta_{\min}$. As $\ln(1/(1-\gamma)) \geq \gamma$, it therefore suffices to show that $1 - \alpha \leq \frac{\gamma}{\ln(1/\delta_{\min})}$. But from the statement of the lemma, $1 - \alpha \leq 1/\tau_0 < \frac{\gamma}{\ln(1/\delta_{\min})}$, completing the proof of (74).

Combining (73) and (74), we get that:

$$\sum_{i=1}^{m} y_i(x) \leq \frac{(1+3\varepsilon)(1+2\gamma)}{1-\gamma} \sum_{j=1}^{n} w_j x_j^{1-\alpha} \leq (1+5\varepsilon) \sum_{j=1}^{n} w_j x_j^{1-\alpha}, \tag{75}$$

where the second inequality follows from $\varepsilon \leq 1/6$, $\gamma = \varepsilon/4$.

Using (75), we can bound the duality gap (Eq. (72)) as:

$$G_\alpha(x, y(x)) \leq \sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{1-\alpha} \left(\alpha \left(\frac{x_j^\alpha \sum_{i=1}^{m} y_i A_{ij}}{w_j} \right)^{\frac{\alpha-1}{\alpha}} - 1 + (1-\alpha)(1+5\varepsilon) \right). \tag{76}$$

To complete the proof, recall from Lemma 4.19 that in any round of the algorithm, for all j: $\frac{x_j \sum_{i=1}^{m} y_i(x) A_{ij}}{w_j} \geq (1-\gamma)^{\tau_0}$. As $\alpha < 1$ and $x_j \in [0, 1]$, $\forall j$, it holds that $x_j^\alpha \geq x_j$, $\forall j$, and therefore:

$$\frac{x_j^\alpha \sum_{i=1}^{m} y_i(x) A_{ij}}{w_j} \geq (1-\gamma)^{\tau_0}, \quad \forall j. \tag{77}$$

Finally, recalling that $1 - \alpha \leq 1/\tau_0$, and combining (77) with (76), we get:

$$\begin{aligned}
G_\alpha(x, y(x)) &\leq \sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{1-\alpha} \left(\alpha \left(\frac{1}{1-\gamma} \right)^{1/\alpha} - 1 + (1-\alpha)(1+5\varepsilon) \right) \\
&\leq \sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{1-\alpha} ((1+2\gamma)^{1/\alpha} - 1 + (1-\alpha)(1+5\varepsilon)) \\
&\leq \sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{1-\alpha} (1 + \varepsilon - 1 + (1-\alpha)(1+5\varepsilon)) \\
&\leq 2\varepsilon \sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{1-\alpha},
\end{aligned}$$

where the third inequality follows from $\alpha \geq 1 - 1/\tau_0 \geq 1 - \frac{\gamma\varepsilon}{5} \geq 1 - \frac{\varepsilon^2}{20}$, and the fourth inequality follows from $1 - \alpha < \varepsilon/2$ and $\varepsilon \leq 1/6$. $\qquad\square$

Lemma 4.32. *Let x be an ε−approximate solution to a 1-fair packing problem (P_1) returned by α-FAIRPSOLVER. Then, for any $\alpha \in (1, 1 + 1/\tau_0]$, where $\tau_0 = \frac{1}{\beta} \ln(\frac{1}{\delta_{\min}})$, x is also a 2ε−approximate solution to (P_α), where the only difference between (P_1) and (P_α) is in the value of α in the objective.*

Proof. Similar to the proof of Lemma 4.31, we will fix an x from some stationary round of α-FAIRPSOLVER run on (P_1), and argue that the same x 2ε−approximates (P_α) by bounding the duality gap $G_\alpha(x, y(x))$, although we will need to use a different set of inequalities since now $\alpha > 1$. Similar to the proof of Lemma 4.31, as x is (primal-)feasible for (P_1), x and $y(x)$ are primal- and dual-feasible for (P_α).

By the same token as in the proof of Lemma 4.31:

$$\sum_{i=1}^{m} y_i(x) \leq (1+3\varepsilon)(1+2\gamma) \sum_{j=1}^{n} w_j.$$

As $\alpha > 1$ and $x_j \in (0, 1]$, $\forall j$, we have that $x_j^{1-\alpha} \geq 1$, $\forall j$, and therefore:

$$\sum_{i=1}^{m} y_i(x) \leq (1+3\varepsilon)(1+2\gamma) \sum_{j=1}^{n} w_j x_j^{1-\alpha} \leq (1+4\varepsilon) \sum_{j=1}^{n} w_j x_j^{1-\alpha}. \tag{78}$$

Therefore, we can write for the duality gap:

$$G_\alpha(x, y(x)) \leq \sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{1-\alpha} \left(\alpha \left(\frac{x_j^{\alpha} \sum_{i=1}^{m} y_i A_{ij}}{w_j} \right)^{\frac{\alpha-1}{\alpha}} - 1 + (1-\alpha)(1+4\varepsilon) \right) \tag{79}$$

$$= -\sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{1-\alpha} \left(-\alpha \left(\frac{x_j^{\alpha} \sum_{i=1}^{m} y_i A_{ij}}{w_j} \right)^{\frac{\alpha-1}{\alpha}} + 1 + (\alpha-1)(1+4\varepsilon) \right). \tag{80}$$

Notice that, as $\alpha > 1$, the objective for (P_α), $\sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{1-\alpha}$, is now negative.

Using the same arguments as in the proof of Lemma 4.31, it is straightforward to show that $x_j^{\alpha-1} \geq 1 - \gamma, \forall j$. From Lemma 4.19, we have that $\frac{x_j \sum_i y_i(x) A_{ij}}{w_j} \geq (1-\gamma)^{\tau_0}, \forall j$, and therefore:

$$\frac{x_j^{\alpha} \sum_{i=1}^{m} y_i(x) A_{ij}}{w_j} = \frac{x_j^{1-\alpha} \cdot x_j \sum_{i=1}^{m} y_i(x) A_{ij}}{w_j}$$

$$\geq (1-\gamma)^{\tau_0+1}. \tag{81}$$

Recalling that $\alpha - 1 \leq 1/\tau_0$ (by the statement of the lemma) and using (81), we have:

$$\left(\frac{x_j^{\alpha} \sum_{i=1}^{m} y_i A_{ij}}{w_j} \right)^{\frac{\alpha-1}{\alpha}} \geq (1-\gamma)^{(\tau_0+1)/(\tau_0(1+1/\tau_0))}$$

$$= (1-\gamma). \tag{82}$$

Finally, plugging (82) into (80), we have:

$$G_\alpha(x, y(x)) \leq -\sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{1-\alpha} \left(-\alpha(1-\gamma) + 1 + (\alpha-1)(1+4\varepsilon) \right)$$

$$= -\sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{1-\alpha} \left(\alpha \cdot \frac{1}{4}\varepsilon + 4\varepsilon(\alpha-1) \right)$$

$$\leq -\varepsilon \sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{1-\alpha}, \tag{83}$$

where the equality follows from $\gamma = \frac{\varepsilon}{4}$, and the last inequality follows from $\alpha - 1 \leq \frac{1}{\tau_0} < \frac{\varepsilon}{20}$. $\square$

Finally, we consider the asymptotics of $\alpha-$fair allocations, as α becomes large. This result complements the result from [38] that states that $\alpha-$fair allocations approach the max-min fair one as $\alpha \to \infty$ by showing how fast the max-min fair allocation is reached as a function of α, R_w, n, and $A_{\max}$. First, for completeness, we provide the definition of max-min fairness.

Definition 4.33. *(Max-min fairness [10].) Let $\mathcal{R} \subset \mathbb{R}_+^n$ be a compact and convex set. A vector $x \in \mathcal{R}$ is max-min fair on $\mathcal{R}$ if for any vector $z \in \mathcal{R}$ it holds that: if for some $j \in \{1, ..., n\}$ $z_j > x_j$, then there exists $k \in \{1, ..., n\}$ such that $z_k < x_k$ and $x_k \leq x_j$.*

On a compact and convex set $\mathcal{R} \subset \mathbb{R}^n$, the max-min fair vector is unique (see, e.g., [44, 45]). The following lemma shows that for $\alpha \geq \varepsilon^{-1} \ln(R_w n A_{\max})$, the $\alpha-$fair vector and the max-min fair vector are $\varepsilon-$close to each other. Notice that because of a very large gradient of $p_\alpha(x)$ as α becomes large, the max-min fair solution provides only an $O(\varepsilon\alpha)-$approximation to (P_α).

Lemma 4.34. *Let x^* be the optimal solution to $(P_\alpha) = \max\{p_\alpha(x) : Ax \leq 1, x \geq 0\}$, z^* be the max-min fair solution for the convex and compact set determined by the constraints from (P_α). Then if $\alpha \geq \varepsilon^{-1} \ln(R_w n A_{\max})$, we have that:*

1. *$p_\alpha(x^*) \leq (1 - \varepsilon(\alpha-1))p_\alpha(z^*)$, i.e., z^* is an $\varepsilon(\alpha-1)-$approximate solution to (P_α), and*

2. *$(1-\varepsilon)z_j^* \leq x_j^* \leq (1+\varepsilon)z_j^*$, for all $j \in \{1, ..., n\}$.*

41

Proof. Suppose that, starting with z^*, we want to construct a solution z that is feasible in (P_α) and is such that $p_\alpha(z) > p_\alpha(z^*)$. Then we need to increase at least one coordinate j of z^*. Suppose that we increase a coordinate j by a factor $1 + \varepsilon$, so that $z_j = (1 + \varepsilon)z_j^*$. Since z^* is the max-min fair vector, to keep z feasible, the increase over the jth coordinate must be at the expense of decreasing some other coordinates k that satisfy $z_k^* \leq z_j^*$. We will assume that whenever we decrease the coordinates to keep the solution feasible, we keep the solution Pareto optimal (i.e., we decrease the selected coordinates by a minimum amount). Using Fact 4.8, we have:

$$p_\alpha(z) - p_\alpha(z^*) \leq \sum_{l=1}^n w_l \frac{z_l - z_l^*}{(z_l^*)^\alpha} < w_j \frac{z_j - z_j^*}{(z_j^*)^\alpha} = \varepsilon \cdot w_j (z_j^*)^{1-\alpha}. \tag{84}$$

Now, suppose that we want to further increase the jth coordinate by some small δ. Call that new solution z^1. Then, the total amount by which other coordinates must decrease to keep the solution feasible is at least $\frac{\delta}{A_{\max}}$, since the feasible region is determined by packing constraints and it must be $Az \leq 1$, where $1 \leq A_{ij} \leq A_{\max}, \forall i, j$. Moreover, since z^* is max-min fair, each coordinate k that gets decreased must satisfy $z_k^* \leq z_j^*$. It follows that:

$$\begin{aligned}
p(z^1) - p(z) &\leq \sum_{l=1}^n w_l \frac{z_l^1 - z_l}{(z_l)^\alpha} \\
&= w_j \frac{\delta}{(1+\varepsilon)^\alpha (z_j^*)^\alpha} + \sum_{k: z_k^1 < z_k} w_k \frac{z_k^1 - z_k}{(z_k)^\alpha} \\
&\leq w_{\max} \frac{\delta}{(1+\varepsilon)^\alpha (z_j^*)^\alpha} - w_{\min} \frac{\delta/A_{\max}}{(z_j^*)^\alpha} \\
&= \frac{\delta(w_{\max} - (1+\varepsilon)^\alpha w_{\min}/A_{\max})}{(1+\varepsilon)^\alpha (z_j^*)^\alpha} \\
&\leq 0. \tag{85}
\end{aligned}$$

The last inequality can be verified by solving the inequality $w_{\max} - (1+\varepsilon)^\alpha w_{\min}/A_{\max} \leq 0$ for α, and verifying that it is implied by the initial assumption that $\alpha \geq \varepsilon^{-1} \ln(R_w n A_{\max})$.

Therefore, the maximum amount by which any coordinate of z^* can be increased to improve the value of the objective $p_\alpha(.)$ is by a multiplicative factor of at most $(1 + \varepsilon)$. Since we can construct x^*, the optimal solution to (P_α), starting with z^* and by choosing a set of coordinates j that we want to increase and by only decreasing coordinates k such that $z_k^* \leq z_j^*$ whenever coordinate j is increased, it follows that $x_j^* \leq (1 + \varepsilon)z_j^*, \forall j$.

Moreover, from (84) and (85):

$$p_\alpha(z^1) - p_\alpha(z^*) = p(z^1) - p(z) + p(z) - p(z^*) < \varepsilon \cdot w_j (z_j^*)^{1-\alpha},$$

and we can conclude that:

$$p_\alpha(x^*) - p_\alpha(z^*) < \sum_{j=1}^n \varepsilon \cdot w_j (z_j^*)^{1-\alpha} = \varepsilon(1-\alpha) \cdot p_\alpha(z^*),$$

which means that z^* is an $\varepsilon(\alpha - 1)-$approximate solution to (P_α).

Now consider the coordinates we need to decrease when we construct a solution z from z^*, such that $p_\alpha(z) > p_\alpha(z^*)$. Suppose that to increase some other coordinates, a coordinate k is decreased by a factor $(1-\varepsilon)$: $z_k = (1-\varepsilon)z_k^*$. As z^* is max-min fair, only coordinates larger than z_k^* can increase at the expense of decreasing z_k^*. Suppose now that we decrease the kth coordinate further by some small δ. Call that solution z^1. Then the maximum number of other coordinates j that can further increase is $\min\{n - 1, m\} < n$. Moreover, each coordinate j that gets increased satisfies $z_j^* \geq z_k^*$, and can be increased by at most $A_{\max}\delta$.

Using Fact 4.8, it follows that:

$$
\begin{aligned}
p_\alpha(z^1) - p_\alpha(z) &\leq \sum_{l=1}^{n} w_l \frac{z_l^1 - z_l}{(z_l)^\alpha} \\
&= -w_k \frac{\delta}{(1-\varepsilon)(z_k^*)^\alpha} + \sum_{j:z_j^1 > z_j} w_j \frac{z_j^1 - z_j}{(z_j^*)^\alpha} \\
&< -w_{\min} \frac{\delta}{(1-\varepsilon)(z_k^*)^\alpha} + n w_{\max} \frac{A_{\max}\delta}{(z_k^*)^\alpha} \\
&= \frac{\delta(n w_{\max} A_{\max}(1-\varepsilon)^\alpha - w_{\min})}{(z_k^*)^\alpha} \\
&\leq 0,
\end{aligned}
\tag{86}
$$

where the last inequality follows from $(1-\varepsilon)^\alpha \leq (R_w n A_{\max})^{-1}$, which is implied by the initial assumption that $\alpha \geq \varepsilon^{-1}\ln(R_w n A_{\max})$.

Therefore, using (86), the kth coordinate can decrease by at most a multiplicative factor $(1-\varepsilon)$. Using similar arguments as for increasing the coordinates, it follows that $x_j^* \geq (1-\varepsilon)z_j^*, \forall j$. □

5   Conclusion

We presented an efficient stateless distributed algorithm for the class of α-fair packing problems. To the best of our knowledge, this is the first algorithm with poly-logarithmic convergence time in the input size. Additionally, we obtained results that characterize the fairness and asymptotic behavior of allocations in weighted $\alpha-$fair packing problems that may be of independent interest. An interesting open problem is to determine the class of objective functions for which the presented techniques yield fast and stateless distributed algorithms, together with a unified convergence analysis. This problem is especially important in light of the fact that α-fair objectives are not Lipschitz continuous, do not have a Lipschitz gradient, and their dual gradient's Lipschitz constant scales at least linearly with n and $A_{\max}$. Therefore, the properties typically used in fast first-order methods are lacking [2, 40]. Finally, for applications of α-fair packing that do not require uncoordinated updates, it seems plausible that the dependence on ε^{-1} in the convergence bound can be improved from ε^{-5} to ε^{-3} by relaxing the requirement for asynchronous updates, similarly as was done in [3] over [7].

References

[1] Z. Allen-Zhu and L. Orecchia. Nearly-linear time positive LP solver with faster convergence rate. In *Proc. ACM STOC'15*, 2015.

[2] Z. Allen-Zhu and L. Orecchia. A novel, simple interpretation of Nesterov's accelerated method as a combination of gradient and mirror descent, Jan. 2015. arXiv preprint, `http://arxiv.org/abs/1407.1537`.

[3] Z. Allen-Zhu and L. Orecchia. Using optimization to break the epsilon barrier: A faster and simpler width-independent algorithm for solving positive linear programs in parallel. In *Proc. ACM-SIAM SODA'15*, 2015.

[4] A. B. Atkinson. On the measurement of inequality. *Journal of Economic Theory*, 2(3):244–263, 1970.

[5] B. Awerbuch, Y. Azar, and R. Khandekar. Fast load balancing via bounded best response. In *Proc. ACM-SIAM SODA'08*, 2008.

[6] B. Awerbuch and R. Khandekar. Greedy distributed optimization of multi-commodity flows. In *Proc. ACM PODC'07*, 2007.

[7] B. Awerbuch and R. Khandekar. Stateless distributed gradient descent for positive linear programs. *SIAM Journal on Computing*, 38(6):2468–2486, 2009.

[8] Y. Bartal, J. Byers, and D. Raz. Global optimization using local information with applications to flow control. In *Proc. IEEE FOCS'97*, 1997.

[9] A. Beck, A. Nedic, A. Ozdaglar, and M. Teboulle. An $O(1/k)$ gradient method for network resource allocation problems. *IEEE Transactions on Control of Network Systems*, 1(1):64–73, 2014.

[10] D. Bertsekas and R. Gallager. *Data Networks*. Prentice Hall, 1992.

[11] D. Bertsimas, V. F. Farias, and N. Trichakis. On the efficiency-fairness trade-off. *Management Science*, 58(12):2234–2250, 2012.

[12] T. Bonald and J. Roberts. Multi-resource fairness: Objectives, algorithms and performance. In *Proc. ACM SIGMETRICS'15*, 2015.

[13] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2009.

[14] A. Charny, D. D. Clark, and R. Jain. Congestion control with explicit rate indication. In *Proc. IEEE ICC'95*, 1995.

[15] Y. K. Cheung, R. Cole, and N. Devanur. Tatonnement beyond gross substitutes?: Gradient descent to the rescue. In *Proc. ACM STOC'13*, 2013.

[16] S. Dolev. *Self-stabilization*. MIT press, 2000.

[17] L. Fleischer. Approximating fractional multicommodity flow independent of the number of commodities. *SIAM Journal on Discrete Mathematics*, 13(4):505–520, 2000.

[18] N. Garg and J. Könemann. Faster and simpler algorithms for multicommodity flow and other fractional packing problems. *SIAM Journal on Computing*, 37(2):630–652, 2007.

[19] N. Garg and N. Young. On-line end-to-end congestion control. In *Proc. IEEE FOCS'02*, 2002.

[20] A. Ghodsi, M. Zaharia, B. Hindman, A. Konwinski, S. Shenker, and I. Stoica. Dominant resource fairness: Fair allocation of multiple resource types. In *Proc. USENIX NSDI'11*, 2011.

[21] S. Im, J. Kulkarni, and K. Munagala. Competitive algorithms from competitive equilibria: Non-clairvoyant scheduling under polyhedral constraints. In *Proc. ACM STOC'14*, 2014.

[22] J. Jaffe. Bottleneck flow control. *IEEE Transactions on Communications*, 29(7):954–962, 1981.

[23] K. Jain and V. Vazirani. Eisenberg-gale markets: Algorithms and structural properties. In *Proc. ACM STOC'07*, 2007.

[24] C. Joe-Wong, S. Sen, T. Lan, and M. Chiang. Multiresource allocation: Fairness-efficiency tradeoffs in a unifying framework. *IEEE/ACM Transactions on Networking*, 21(6):1785–1798, 2013.

[25] F. Kelly, A. Maulloo, and D. Tan. Rate control for communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research Society*, 49(3):237–252, 1998.

[26] F. Kelly and E. Yudovina. *Stochastic networks*, volume 2. Cambridge University Press, 2014.

[27] J. Kleinberg, Y. Rabani, and É. Tardos. Fairness in routing and load balancing. In *Proc. IEEE FOCS'99*, 1999.

[28] C. Koufogiannakis and N. Young. Beating simplex for fractional packing and covering linear programs. In *Proc. IEEE FOCS'07*, 2007.

[29] F. Kuhn, T. Moscibroda, and R. Wattenhofer. The price of being near-sighted. In *Proc. ACM-SIAM SODA'06*, 2006.

[30] A. Kumar and J. Kleinberg. Fairness measures for resource allocation. In *Proc. IEEE FOCS'00*, 2000.

[31] T. Lan, D. Kao, M. Chiang, and A. Sabharwal. An axiomatic theory of fairness in network resource allocation. In *Proc. IEEE INFOCOM'10*, 2010.

[32] S. Low, F. Paganini, and J. Doyle. Internet congestion control. *IEEE Control Systems*, 22(1):28–43, 2002.

[33] M. Luby and N. Nisan. A parallel approximation algorithm for positive linear programming. In *Proc. ACM STOC'93*, 1993.

[34] J. Marašević, C. Stein, and G. Zussman. Max-min fair rate allocation and routing in energy harvesting networks: Algorithmic analysis. In *Proc. ACM MobiHoc'14*, 2014.

[35] B. McCormick, F. Kelly, P. Plante, P. Gunning, and P. Ashwood-Smith. Real time alpha-fairness based traffic engineering. In *Proc. ACM HotSDN'14*, 2014.

[36] N. Megiddo. Optimal flows in networks with multiple sources and sinks. *Mathematical Programming*, 7(1):97–107, 1974.

[37] D. Mitrinovic and P. Vasic. *Analytic inequalities*, volume 1. Springer, 1970.

[38] J. Mo and J. Walrand. Fair end-to-end window-based congestion control. *IEEE/ACM Transactions on Networking*, 8(5):556–567, Oct. 2000.

[39] D. Mosk-Aoyama, T. Roughgarden, and D. Shah. Fully distributed algorithms for convex optimization problems. In *Proc. DISC'07*, 2007.

[40] Y. Nesterov. *Introductory lectures on convex optimization*, volume 87. Springer Science & Business Media, 2004.

[41] F. Paganini, Z. Wang, J. Doyle, and S. Low. Congestion control for high performance, stability, and fairness in general networks. *IEEE/ACM Transactions on Networking*, 13(1):43–56, 2005.

[42] C. Papadimitriou and M. Yannakakis. Linear programming without the matrix. In *Proc. ACM STOC'93*, 1993.

[43] S. Plotkin, D. Shmoys, and É. Tardos. Fast approximation algorithms for fractional packing and covering problems. *Mathematics of Operations Research*, 20(2):257–301, 1995.

[44] B. Radunović and J.-Y. L. Boudec. A unified framework for max-min and min-max fairness with applications. *IEEE/ACM Transactions on Networking*, 15(5):1073–1083, 2007.

[45] S. Sarkar and L. Tassiulas. Fair allocation of discrete bandwidth layers in multicast networks. In *Proc. IEEE INFOCOM'00*, 2000.

[46] Y. Yi and M. Chiang. Stochastic network utility maximisation - a tribute to Kelly's paper published in this journal a decade ago. *European Transactions on Telecommunications*, 19(4):421–442, 2008.

[47] N. Young. Sequential and parallel algorithms for mixed packing and covering. In *Proc. IEEE FOCS'01*, 2001.

A Scaling Preserves Approximation

Let the α-fair allocation problem be given in the form:

$$(Q_\alpha) \quad \mathbf{max}\Big\{ \sum_{j=1}^{n} w_j f_\alpha(x_j) : Ax \leq b, x \geq 0 \Big\}, \text{ where } f_\alpha(x_j) = \begin{cases} \ln(x_j), & \text{if } \alpha = 1 \\ \dfrac{x_j^{1-\alpha}}{1-\alpha}, & \text{if } \alpha \neq 1 \end{cases},$$

w is an $n-$length vector of positive weights, x is the vector of variables, A is an $n \times m$ constraint matrix, and b is an $m-$length vector with positive entries. Denote $p_\alpha(x) = \sum_{j=1}^{n} w_j f_\alpha(x_j)$.

It is not hard to see that the assumption $b_i = 1 \ \forall i$ is without loss of generality, since for $b_i \neq 1$ we can always divide both sides of the inequality by b_i and obtain 1 on the right-hand side, since for (non-trivial) packing problems $b_i > 0$. Therefore, we can assume that the input problem has constraints of the form $A \cdot x \leq \mathbb{1}$, although it may not necessarily be the case that $A_{ij} \geq 1 \ \forall A_{ij} \neq 0$.

The remaining transformation that is performed on the input problem is:

$$\widehat{x}_j = c \cdot x_j, \quad \widehat{A}_{ij} = A_{ij}/c.$$

where

$$c = \begin{cases} \min_{i,j:A_{ij}\neq 0} A_{ij}, & \text{if } \min_{i,j:A_{ij}\neq 0} A_{ij} < 1 \\ 1, & \text{otherwise} \end{cases}.$$

The problem (Q_α) after the scaling becomes:

$$\begin{aligned} \mathbf{max} \quad & \sum_{j=1}^{n} w_j f_\alpha(\widehat{x}_j) \cdot c^{1-\alpha} \\ \mathbf{s.t.} \quad & \widehat{A}\widehat{x} \leq \mathbb{1} \\ & \widehat{x} \geq 0 \end{aligned} \qquad \Leftrightarrow \qquad \begin{aligned} (P_\alpha) \quad \mathbf{max} \quad & \sum_{j=1}^{n} w_j f_\alpha(\widehat{x}_j) \\ \mathbf{s.t.} \quad & \widehat{A}\widehat{x} \leq \mathbb{1} \\ & \widehat{x} \geq 0, \end{aligned}$$

as $c^{1-\alpha}$ is a positive constant. Recall that α-FAIRPSOLVER returns an approximate solution to (P_α), and observe that x is feasible for (Q_α) if and only if $\widehat{x}$ is feasible for (P_α).

Choose the dual variables (Lagrange multipliers) for the original problem (Q_α) as:

$$y_i = c^{\alpha-1} C \cdot e^{\kappa(\sum_{i=1}^{n} A_{ij}x_j - 1)} = c^{\alpha-1} C \cdot e^{\kappa(\sum_{i=1}^{n} \widehat{A}_{ij}\widehat{x}_j - 1)} = c^{\alpha-1}\widehat{y}_i, \tag{87}$$

and notice that

$$x_j{}^\alpha \sum_{i=1}^{m} y_i A_{ij} = \widehat{x}_j^\alpha \cdot c^{-\alpha} \cdot \sum_{i=1}^{m} (c^{\alpha-1} \cdot \widehat{y}_i \cdot c \cdot \widehat{A}_{ij}) = \widehat{x}_j^\alpha \sum_{i=1}^{m} \widehat{y}_i \widehat{A}_{ij}. \tag{88}$$

It is clear that y_i's are feasible dual solutions, since the only requirement for the duals is non-negativity.

A.1 Approximation for Proportional Fairness

Recall (from (2)) that the duality gap for a given primal- and dual-feasible x and y is given as:

$$G(x,y) = \sum_{j=1}^{n} w_j \ln(w_j) - \sum_{j=1}^{n} w_j \ln\left(x_j \sum_{i=1}^{m} y_i A_{ij} \right) + \sum_{i=1}^{m} y_i - 1.$$

Since $\alpha = 1$, we have that $\widehat{y}_i = y_i$ for all i, and using (88), it follows that

$$G(\widehat{x}, \widehat{y}) = G(x,y).$$

Since we demonstrate an additive approximation for the proportional fairness via the duality gap: $p(\widehat{x}^*) - p(\widehat{x}) \leq G(\widehat{x}, \widehat{y})$, the same additive approximation follows for the original (non-scaled) problem.

A.2 Approximation for α-Fairness and $\alpha \neq 1$

For $\alpha \neq 1$, we show that the algorithm achieves a multiplicative approximation for the scaled problem. In particular, we show that after the algorithm converges we have that: $p_\alpha(\widehat{x}^*) - p_\alpha(\widehat{x}) \leq r_\alpha p_\alpha(\widehat{x})$, where $\widehat{x}^*$

is the optimal solution, $\widehat{x}$ is the solution returned by the algorithm, and r_α is a constant.

Observe that since $\widehat{x} = c \cdot x$, we have that $p_\alpha(\widehat{x}^*) = c^{1-\alpha}p(x^*)$ and $p_\alpha(\widehat{x}) = c^{1-\alpha}p_\alpha(x)$. Therefore:

$$p_\alpha(x^*) - p_\alpha(x) = c^{\alpha-1}(p_\alpha(\widehat{x}^*) - p_\alpha(\widehat{x}))$$
$$\leq c^{\alpha-1} \cdot r_\alpha p_\alpha(\widehat{x})$$
$$= r_\alpha p_\alpha(x).$$

B Primal, Dual, and the Duality Gap

B.1 Proportionally Fair Resource Allocation

In this section we consider $(w, 1)$-proportional resource allocation, often referred to as the weighted proportionally fair resource allocation. Recall that the primal is of the form:

$$(P_1) \quad \textbf{max} \quad \sum_{j=1}^n w_j \ln(x_j)$$
$$\textbf{s.t.} \quad Ax \leq \mathbb{1},$$
$$x \geq 0.$$

The Lagrangian for this problem can be written as:

$$L_1(x; y, z) = \sum_{j=1}^n w_j \ln(x_j) + \sum_{i=1}^m y_i \cdot \left(1 - \sum_{j=1}^n A_{ij}x_j - z_i\right),$$

where $y_1, ..., y_m$ are Lagrange multipliers, and $z_1, ..., z_m$ are slack variables. The dual to this problem is:

$$(D_1) \quad \textbf{min} \quad g(y)$$
$$\textbf{s.t.} \quad y \geq 0,$$

where $g(y) = \max_{x,z \geq 0} L(x; y, z)$. To maximize $L_1(x; y, z)$, we first differentiate with respect to x_j, $j \in \{1, ..., n\}$:

$$\frac{\partial L_1(x; y, z)}{\partial x_j} = \frac{w_j}{x_j} - \sum_{i=1}^m y_i A_{ij} = 0,$$

which gives:

$$x_j \cdot \sum_{i=1}^m y_i A_{ij} = w_j, \quad \forall j \in \{1, ..., n\}. \tag{89}$$

Plugging this back into the expression for $L_1(x; y, z)$, and noticing that, since $y_i, z_i \geq 0 \ \forall i \in \{1, ..., m\}$, $L_1(x; y, z)$ is maximized for $z_i = 0$, we get that:

$$g_1(y) = \sum_{j=1}^n w_j \ln\left(\frac{w_j}{\sum_{i=1}^m y_i A_{ij}}\right) + \sum_{i=1}^m y_i - \sum_{i=1}^m y_i \sum_{j=1}^n \frac{A_{ij}w_j}{\sum_{k=1}^m y_k A_{kj}}$$

$$= \sum_{j=1}^n w_j \ln(w_j) - \sum_{j=1}^n w_j \ln\left(\sum_{i=1}^m y_i A_{ij}\right) + \sum_{i=1}^m y_i - \sum_{j=1}^n w_j \sum_{i=1}^m \frac{y_i A_{ij}}{\sum_{k=1}^m y_k A_{kj}}$$

$$= \sum_{j=1}^n w_j \ln(w_j) - \sum_{j=1}^n w_j \ln\left(\sum_{i=1}^m y_i A_{ij}\right) + \sum_{i=1}^m y_i - W,$$

since $\sum_{i=1}^m \frac{y_i A_{ij}}{\sum_{k=1}^m y_k A_{kj}} = 1 \ \forall j \in \{1, ..., n\}$, and $\sum_{j=1}^n w_j = W$.

Let $p_1(x) = \sum_{j=1}^n w_j \ln(x_j)$ denote the primal objective. The duality gap for any pair of primal-feasible

x and dual-feasible (nonnegative) y is given by:

$$G_1(x, y) = g_1(y) - p_1(x)$$

$$= -\sum_{j=1}^{n} w_j \ln\left(\frac{x_j \sum_{i=1}^{m} y_i A_{ij}}{w_j}\right) + \sum_{i=1}^{m} y_i - W.$$

Since the primal problem maximizes a concave function over a polytope, the strong duality holds [13], and therefore $G_1(x, y) \geq 0$ for any pair of primal-feasible x and dual-feasible y, with equality if and only if x and y are primal- and dual- optimal, respectively.

B.2 α-Fair Resource Allocation for $\alpha \neq 1$

Recall that for $\alpha \neq 1$ the primal problem is:

$$(P_\alpha) \quad \textbf{max} \quad \sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{1-\alpha} \equiv p_\alpha(x)$$

$$\textbf{s.t.} \quad Ax \leq 1,$$

$$x \geq 0.$$

The Lagrangian for this problem can be written as:

$$L_\alpha(x; y, z) = \sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{1-\alpha} + \sum_{i=1}^{m} y_i \left(1 - \sum_{j=1}^{n} A_{ij} x_j - z_i\right),$$

where y_i and z_i, for $i \in \{1, ..., m\}$, are Lagrangian multipliers and slack variables, respectively.

The dual to (P_α) can be written as:

$$(D_\alpha) \quad \textbf{min} \quad g(y)$$

$$\textbf{s.t.} \quad y \geq 0,$$

where $g_\alpha(y) = \max_{x, z \geq 0} L_\alpha(x; y, z)$.

Since $L_\alpha(x; y, z)$ is differentiable with respect to x_j for $j \in \{1, ..., n\}$, it is maximized for:

$$\frac{\partial L_\alpha(x; y, z)}{\partial x_j} = \frac{w_j}{x_j^\alpha} - \sum_{i=1}^{m} y_i A_{ij} = 0$$

$$\Rightarrow w_j = x_j^\alpha \sum_{i=1}^{m} y_i A_{ij}. \tag{90}$$

As $z_i \cdot y_i \geq 0 \; \forall i \in \{1, ..., m\}$, we get that:

$$g_\alpha(y) = \sum_{j=1}^{n} \frac{w_j}{1-\alpha} \left(\frac{w_j}{\sum_{i=1}^{m} y_i A_{ij}}\right)^{\frac{1-\alpha}{\alpha}} + \sum_{i=1}^{m} y_i - \sum_{i=1}^{m} y_i \sum_{j=1}^{n} A_{ij} \left(\frac{w_j}{\sum_{k=1}^{m} y_k A_{kj}}\right)^{1/\alpha}$$

$$= \sum_{j=1}^{n} \frac{w_j}{1-\alpha} \left(\frac{w_j}{\sum_{i=1}^{m} y_i A_{ij}}\right)^{\frac{1-\alpha}{\alpha}} + \sum_{i=1}^{m} y_i - \sum_{j=1}^{n} w_j^{1/\alpha} \left(\sum_{k=1}^{m} y_k A_{kj}\right)^{-1/\alpha} \sum_{i=1}^{m} A_{ij} y_i$$

$$= \sum_{j=1}^{n} \frac{w_j}{1-\alpha} \left(\frac{w_j}{\sum_{i=1}^{m} y_i A_{ij}}\right)^{\frac{1-\alpha}{\alpha}} + \sum_{i=1}^{m} y_i - \sum_{j=1}^{n} w_j^{1/\alpha} \left(\sum_{i=1}^{m} A_{ij} y_i\right)^{\frac{\alpha-1}{\alpha}}.$$

Similarly as before, for primal-feasible x and dual-feasible y, the duality gap is given as:

$$G_\alpha(x, y) = g_\alpha(y) - p_\alpha(x)$$

48

$$= \sum_{j=1}^{n} \frac{w_j}{1-\alpha} \left(\frac{w_j}{\sum_{i=1}^{m} y_i A_{ij}} \right)^{\frac{1-\alpha}{\alpha}} + \sum_{i=1}^{m} y_i - \sum_{j=1}^{n} w_j^{1/\alpha} \left(\sum_{i=1}^{m} A_{ij} y_i \right)^{\frac{\alpha-1}{\alpha}} - \sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{1-\alpha}$$

$$= \sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{1-\alpha} \left(\left(\frac{w_j}{x_j{}^{\alpha} \sum_{i=1}^{m} y_i A_{ij}} \right)^{\frac{1-\alpha}{\alpha}} - 1 \right) + \sum_{i=1}^{m} y_i - \sum_{j=1}^{n} w_j^{1/\alpha} \left(\sum_{i=1}^{m} A_{ij} y_i \right)^{\frac{\alpha-1}{\alpha}}.$$

Observing that:

$$w_j^{1/\alpha} \left(\sum_{i=1}^{m} A_{ij} y_i \right)^{\frac{\alpha-1}{\alpha}} = w_j \cdot w_j^{-\frac{\alpha-1}{\alpha}} \cdot x_j^{1-\alpha} \cdot x_j^{\alpha \frac{\alpha-1}{\alpha}} \cdot \left(\sum_{i=1}^{m} A_{ij} y_i \right)^{\frac{\alpha-1}{\alpha}}$$

$$= w_j x_j^{1-\alpha} \cdot \left(\frac{x_j{}^{\alpha} \sum_{i=1}^{m} A_{ij} y_i}{w_j} \right)^{\frac{\alpha-1}{\alpha}},$$

we finally get:

$$G_\alpha(x,y) = \sum_{j=1}^{n} w_j \frac{x_j^{1-\alpha}}{1-\alpha} \left(\left(\frac{x_j{}^{\alpha} \sum_{i=1}^{m} y_i A_{ij}}{w_j} \right)^{\frac{\alpha-1}{\alpha}} - 1 \right) + \sum_{i=1}^{m} y_i - \sum_{j=1}^{n} w_j x_j^{1-\alpha} \cdot \left(\frac{x_j{}^{\alpha} \sum_{i=1}^{m} A_{ij} y_i}{w_j} \right)^{\frac{\alpha-1}{\alpha}}.$$